A STREAMLINED MULTIVARIATE ANALYSIS PLATFORM FOR OPTIMIZING MAMMALIAN CELL CULTURE AND BIOLOGICS PRODUCTION

A Thesis
Presented to
The Academic Faculty

By

Andrew N. Mingee

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
Wallace H. Coulter Department of Biomedical Engineering

Georgia Institute of Technology
Emory University
May 2021

A STREAMLINED MULTIVARIATE ANALYSIS PLATFORM FOR OPTIMIZING MAMMALIAN CELL CULTURE AND BIOLOGICS PRODUCTION

Approved By:

Dr. Sakis Mantalaris
Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Mark Styczynski
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Melissa Kemp
Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Eberhard Voit
Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology

Date Approved: July 29, 2021

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Dr. Sakis Mantalaris for taking a chance on me and having the perfect idea for a project tailored to my interests and professional development. I was consistently challenged and learned a lot more during my time at Georgia Tech than I otherwise would have. Next, I would like to thank Dr. Ana Quiroga for being my mentor and support system during my time in the lab. As our only postdoctoral researcher, she went above and beyond to provide me with meaningful work and the necessary guidance to learn and challenge myself. We did a lot of work but had an equal amount of fun working together. I would also like to thank my parents for financially and emotionally supporting my educational endeavors, as getting this degree would not have been possible without them. Lasty, I'd like to thank my girlfriend for helping to keep me motivated, disciplined, and organized during the completion of this thesis. I would not be the person or student I am today without the support of her, my family, and friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ATP	Adenosine Triphosphate
CD	Chemically Defined
CHO	Chinese Hamster Ovary
CL	Conserved Light
CQA	Critical Quality Attribute
CSV	Comma Separated Values
DAD	Diode Array Detector
DHFR	Dihydrofolate Reductase
DNA	Deoxyribonucleic Acid
DOE	Design of Experiments
ELISA	Enzyme-Linked Immunosorbent Assay
$FADH_2$	Flavin Adenine Dinucleotide
GS	Glutamine Synthetase
GTP	Guanosine Triphosphate
HAT	Hypoxanthine-Aminopterin-Thymidine
HC	Hierarchical Clustering
HCA	Hierarchical Clustering Analysis
HPLC	High Performance Liquid Chromatography
HSC	Hematopoietic Stem Cells
MHC	Major Histocompatibility Complex
MSX	L-Methionine Sulfoximine

MVA	Multivariate Analysis
n	Number of Samples
NADH	Nicotinamide Adenine Dinucleotide
NS0	Non-Secreting Myeloma
p	Data Matrix Row
PBMC	Peripheral Blood Mononuclear Cell
PC	Principal Component
PCA	Principal Component Analysis
PDH	Pyruvate Dehydrogenase
q	Data Matrix Column
QC	Quality Control
ROS	Reactive Oxygen Species
RPM	Rotations Per Minute
S	Covariance Matrix
SHMT	Serine Hydroxymethyltransferase
TCA	Tricarboxylic Acid Cycle
U	Standardized Data Matrix
UV-Vis	Ultraviolet -Visible
VH	Variable Heavy
VL	Variable Light
z	Z-Score

SUMMARY

Biologics-based therapeutics have witnessed remarkable market growth in the past couple of decades with continuing future potential. Expiring patents on tried-and-true monoclonal antibody (mAb) drugs put tremendous pressure on pharmaceutical manufacturing companies to develop biosimilars as quickly and efficiently as possible. One of the biggest bottlenecks in the scale-up of biopharmaceuticals manufacturing is the production of biomolecules, and this is largely a consequence of cell culture constraints. Robust mammalian cell lines, such as Chinese hamster ovary (CHO) or non-secreting myeloma (NS0) cells, have been extensively characterized using -omics technologies; however, there is still more research to be done. One of the most high-impact changes that a pharmaceutical development company can make is the optimization of their cultures, and one of the best ways to do this is through the optimization of media composition and feeding strategy (for fed-batch cultures). Since basal media is often generalized for many cell lines of a particular cell type, supplementations must be made to cater to the specific nutrient requirements for the clone being cultured and biologic being produced. These needs are determined by collecting data on the consumption and production of nutrients and metabolites in the culture media and within the cells. The aim of this project is to culture mAb-producing GS-CHO cells and characterize their metabolic needs and behavior using quantitative methods such as high-performance liquid chromatography (for amino acids, vitamins, and mAb titer), flow cytometry (for cell count and viability), fluorescence measurements (for ATP) and BioProfiling (for glucose, lactate, and inorganic salts). It can be difficult to draw conclusions from so many data, so a streamlined and automated platform for multivariate analysis using

Excel and R programming allows for clearer visualization of the data at a high level using techniques like hierarchical clustering and principal component analysis. This analysis helps elucidate relationships and shifts between the variables, which can be used to improve upon future cultures by informing model-based optimization software and providing a better understanding of specific culture needs.

CHAPTER 1

INTRODUCTION

Bioprocess optimization can be a daunting task due to its complex and costly nature, and for these reasons, many biopharmaceutical companies avoid spending too much time or money on it. Inefficient batch processes still dominate the market, despite theoretically superior alternatives and optimization opportunities [1]. Given the multitude of process parameters that must be considered at each stage of the manufacturing process, this is somewhat understandable. Below is a fishbone diagram recreated from a case study conducted by some of the leading global biopharmaceutical development companies, which visualizes many of the factors of cell culture critical to product development [2].

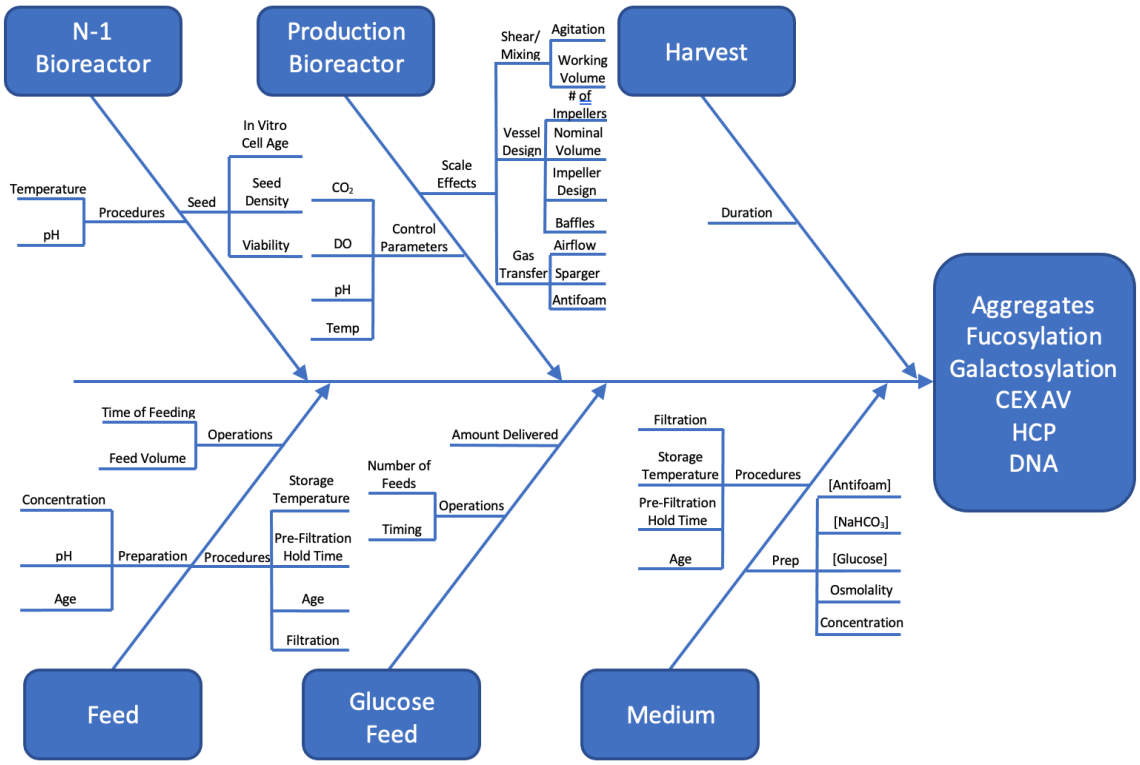


Figure 1.1: Process Parameters for Cell Culture in Biopharmaceutical Manufacturing [2]

The existence of so many process parameters makes it challenging to isolate which ones might be the most impactful when optimized. One strategy to determine the ideal factor conditions and how they influence one another is to employ design of experiments (DOE). This is an optimization strategy that uses applied statistics to generate an array of different experimental conditions that result in the comprehensive testing of multiple process parameters in a highly controlled manner. This systematically evaluates each parameter and its impact on the process responses (such as yield or cell density) to maximize productivity and minimize variability [3]. A technique known as factorial design is used to determine which factors (synonymous to independent variables) are to be varied during experimental runs and how many levels will be tested (low, medium, high, etc.). While DOE-based optimization is successful by design, it often requires that hundreds of runs be conducted, many of which may not be very successful, which is costly in terms of time and resources. For example, using Figure 1.1 as a reference, if even just two levels were to be tested for each of the thirteen factors listed under the production bioreactor process, 8,192 runs would be required to test every combination of levels. Such excessive testing is clearly unrealistic, and there are techniques to decrease the number of required runs given limited resources - such as selectively testing for specific relationships and not others, or choosing to omit the investigation of certain factors or levels entirely based on existing information – but it is easy to imagine how this strategy can be difficult to apply to multi-step industrial bioprocesses [3]. However, this trial-and-error based optimization is one of the most widely utilized techniques by companies, due to its straightforwardness, simplicity and barriers to entry for superior alternatives.

The aim of this thesis is to design a multivariate data analysis platform to inform

bioprocess optimization, with a focus on chemically defined media composition. While there are many offerings on the market for tried-and-true culture media formulations for popular mammalian cell lines, genetic variation between cell line clones and differing chemical compositions between biologics demands specific nutrient requirements which may not be met by generalized culture media formulations [4]. Therefore, this project seeks to profile the metabolic behavior of an industrially relevant mammalian cell line producing a monoclonal antibody by extensively analyzing batch cultures. More specifically, quantitative lab techniques will be used to determine the dynamic concentrations of amino acids, vitamins, metabolites and monoclonal antibody product in spent media samples (collected daily) and develop a streamlined analysis platform to draw conclusions about the cellular metabolism and productivity over time in order to inform a media design-based optimization strategy. This is beneficial because the ability to identify patterns and correlations in a large set of data from a single experiment could significantly reduce the number of further experiments required for optimization by pinpointing potential key nutrients to focus on for supplementation.

CHAPTER 2

BACKGROUND

2.1 What are Monoclonal Antibodies?

Monoclonal antibodies (mAbs) have reshaped the biologics and targeted therapeutics markets since their advent just a few decades ago, and continual improvements in their development yield a promising future [5]. In short, monoclonal antibodies are immunoglobulin G (IgG) proteins secreted by white blood cells that selectively target a single antigen epitope with high specificity. Their incorporation in drug formulations allows for the targeted treatment of a specific disease, greatly reducing side effects and enhancing efficacy. The industrial production of therapeutic mAbs has experienced remarkable growth expansion, having reached a market size of $98 billion in 2017, excluding diagnostic and academic applications [5]. This chapter will provide a brief overview of their historical discovery and characterization, discuss current and future improvements to developmental and manufacturing practices, and explain their diverse applications and clinical relevance. Portions of this chapter are from a forthcoming chapter by myself, Dr. Ana Quiroga and Dr. Sakis Mantalaris in *A Handbook of Molecular Biotechnology*, entitled "Monoclonal Antibodies".

2.1.1 Monoclonal Antibody Structure

Monoclonal antibodies are Y-shaped glycoproteins, and they can be segmented into two regions: the antigen binding fragments (Fab) which make up the two branches of the "Y", and the crystallizable fragment which forms the body of the "Y". The antigen-binding

fragments can be further segmented into their light and heavy chains [6]. Heavy chains have a molecular weight of 50-70 kDa and span the entire length of the antibody molecule [7]. Possessing 4-5 total domains, the heavy chains contain one variable domain (VH) located within the Fab, and the remaining conserved domains (CH) reside in the Fc region. The light chains, in contrast, weigh only 25 kDa and are found exclusively within the antigen-binding fragments [7]. Each light chain possesses one variable domain (VL) and once conserved domain (CL). Disulfide bonds connect the light and heavy chains within the Fabs, and also connect the two heavy chains in the Fc region. The maintenance of these disulfide bonds is critical to the structural and functional integrity of mAbs, so they are considered a critical quality attribute (CQA) of the final pharmaceutical product [8]. All of these structures are collectively depicted in Figure 2.1.

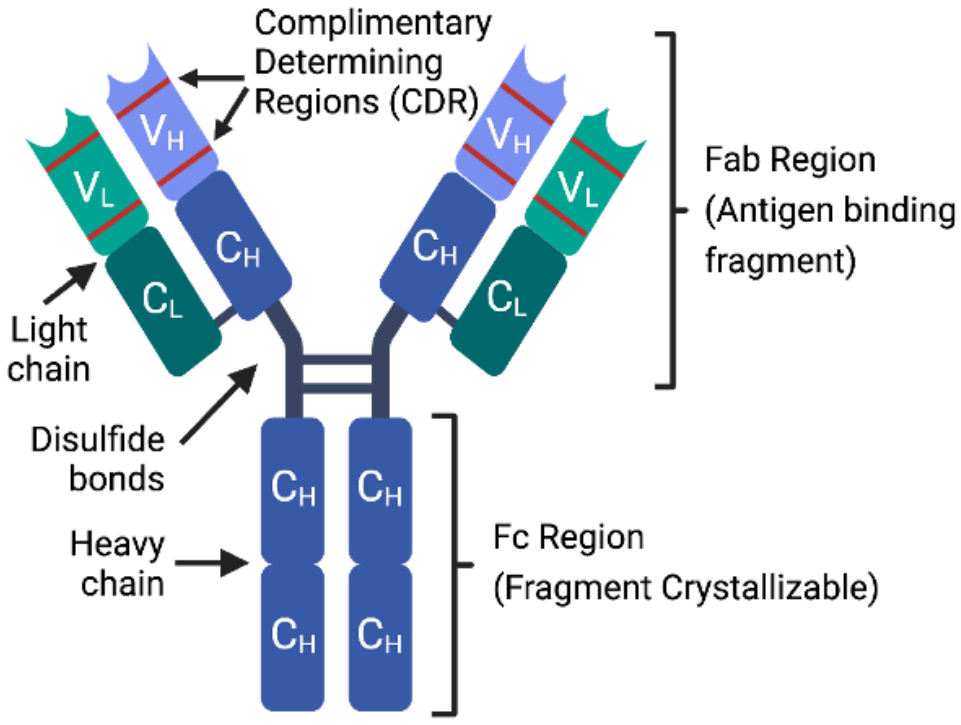


Figure 2.1: Monoclonal Antibody Structure (created with Biorender.com) [6]

2.1.2 Antibody Fragments

Given their larger size, mAbs can be fragmented in numerous different ways to preserve functionality in smaller form factors. These fragments can be engineered to either function independently or can be incorporated into pharmaceutical formulations or delivery methods (such as conjugation to nanoparticles) for targeted therapies. Monoclonal antibody fragments can be broadly categorized as either "true" or "engineered". True fragments describe any type of fragmentation that preserves its structural integrity from the original whole IgG molecule, whereas engineered fragments involve a combination or structural rearrangement of true fragments to make up an entirely new structure which is not represented in the whole molecule. Common examples of both types of fragments are visualized in Figure 2.2.

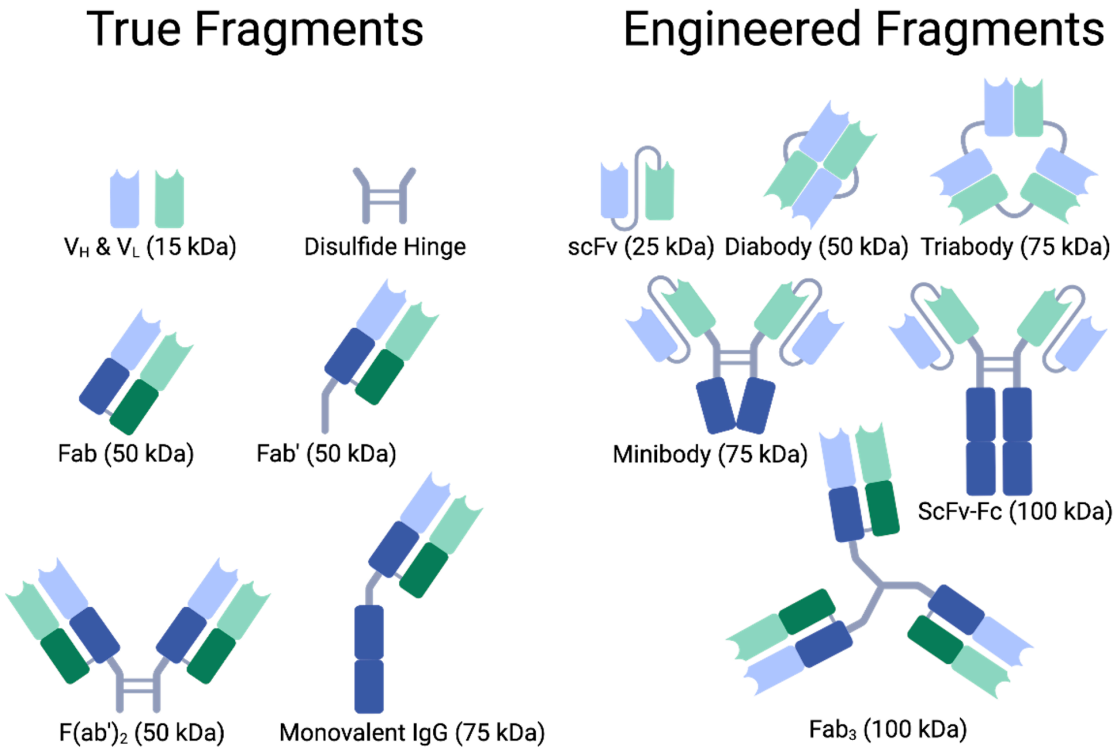


Figure 2.2: Antibody Fragments (created using BioRender.com) [9]

2.1.3 IgG Antibody Glycosylation

Glycosylation, another critical quality attribute of mAbs, describes one of the post-translational modifications to the constant region and is a very important factor in the efficacy and safety of therapeutic mAbs. In short, complex structures of glycosidically-linked monosaccharides (glycans) are joined to either side of the FC region of an IgG antibody after its production, and these glycans must be accurately and consistently reproduced in cultures to maintain the effectiveness and homogeneity of a mAb product. Glycosylation employs complex multi-step biochemical pathways that have been extensively characterized, but the functional purposes of glycosylation are not yet fully understood. It is known that glycosylation in the rough ER serves as a labelling mechanism to identify proper protein folding and also significantly impacts protein-protein interactions by physically modifying the structure of mAbs, preventing unfavorable binding and/or facilitating necessary binding. Glycosylation aides in communication with receptors in the Golgi network which is responsible for facilitating their secretion. The main types of glycosylation are defined by their site of linkage on the F_c region of the mAb: N-linked glycosylation, by far the most common type, involves the binding of the glycan to the amino group of an asparagine residue. O-linked glycosylation entails linkage to the hydroxyl group of a threonine or serine residue and describes approximately 10% of glycosylation patterns. The specific structural composition of a given glycan to a mAb is termed a glycoform, and robust bioprocessing aims to avoid glycoform heterogeneity in the product. A summary of common glycosylation steps and incorporated sugars (such as galactose, mannose and fucose) is given in Figure 2.3.

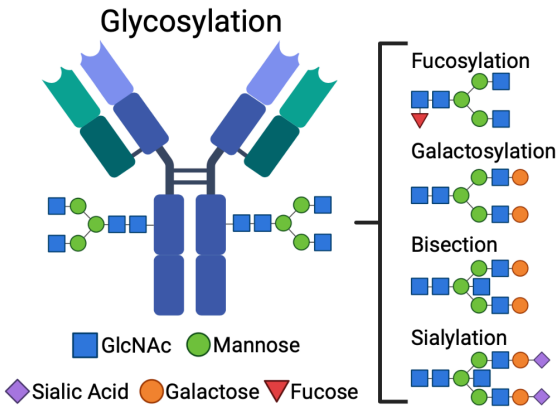


Figure 2.3: IgG Antibody Glycosylation (created using BioRender.com) [10]

2.1.4 Bispecific IgG Antibodies

Monoclonal IgG antibodies, which are innately monospecific, can be engineered to be alternatively "bispecific". As previously described, IgG antibodies have two antigen binding sites, and this pair of Fabs is naturally identical. However, the substitution of an Fab with one derived from a different mAb results in an antibody that yields specific binding to two epitopes of an antigen, or even two separate antigens entirely (Figure 2.4). Monoclonal antibodies that possess heterogenous binding sites are advantageous in situational applications, which will be described in greater detail in a later section.

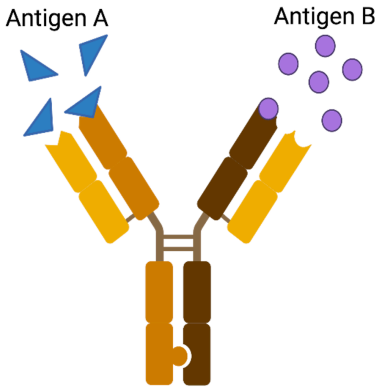


Figure 2.4: Bispecific IgG Antibody (created using BioRender.com)

2.1.5 Humanization of Monoclonal Antibodies

Since monoclonal antibodies are conventionally produced by murine cells, they naturally invoke an unwanted immunogenic response when administered in humans. This is because murine genomes possess enough differences from the human genome that the mAbs produced fundamentally lack the same post-translational modifications found in proteins endogenously produced in humans. Therefore, initial mAb therapies were flagged as foreign molecules by patient immune systems and neutralized before they could effectively enact their therapeutic action, making them largely unsuccessful. In general, they also yielded lower affinity and cytotoxicity than desired. This prompted immediate efforts to "humanize" mAbs and approve efficacy, of which there are multiple approaches (Figure 2.5).

Chimerization of mAbs involves the combination of structural components genetically derived from humans and rodents. More specifically, the variable domains of murine mAbs are combined with the constant region of a human IgG antibody, resulting in a mAb that is genetically 65%-75% human whilst maintaining the specific antigen binding fragments produced by murine hybridomas. Abciximab, a cardiovascular disease therapy targeting the GPIIb/IIIa complex on platelets to prevent aggregation, received FDA approval in 1994 and was the first medication to implement chimeric antibodies [11]. Humanized antibodies are conceptually similar but incorporate an even smaller fragment of murine mAbs. Instead of incorporating the entire variable domain into a human IgG antibody, only the hypervariable sequences responsible specifically for antigen binding are grafted, resulting in a mAb that is genetically more than 90% human. The first humanized mAb to receive FDA approval was Daclizumab which targeted the IL-2 receptor for

prevention of transplant rejection [11]. The production of fully human mAbs can be obtained using transgenic mice hybridomas or phage display technology, which will be described in the following section. Adalimumab, an immunosuppressant primarily for the treatment of rheumatoid arthritis, received FDA approval in 2002 and was the first fully human mAb on the market. Developed by AbbVie under the brand name Humira, it achieved huge success and was the bestselling mAb therapeutic in 2018 [11]. The level of humanization for a clinically approved mAb can be discerned by its suffix nomenclature: "-omab" describes fully murine mAbs, "-ximab" describes chimeric mAbs, "-zumab" describes humanized mAbs, and "-umab" describes fully human mAbs.

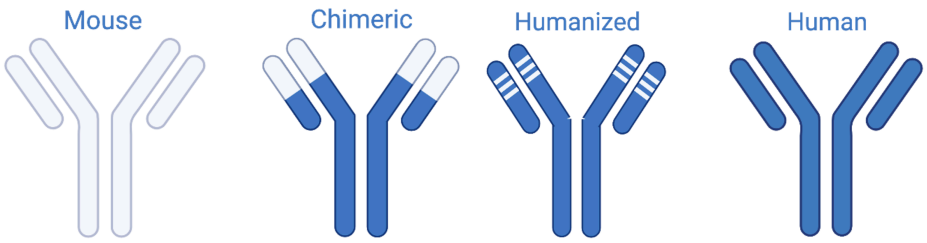


Figure 2.5: Humanization of Monoclonal Antibodies (created using BioRender.com)

2.2 Antibody Discovery

Antibodies found in blood serum are heterogeneous, meaning that they are composed of a polyclonal mixture that recognizes different epitopes of the same antigens and also different antigens entirely. Because B-cells are only capable of recognizing and secreting antibodies for a single epitope of a single antigen, there exists vast heterogeneity within a single person or animal's population of B-cells to provide immunity against a diverse range of diseases. This diversity is made possible by the processes described in 1957 by immunologists Frank Burnet and David Talmage in their "clonal selection theory" [12].

They hypothesized that hematopoietic stem cells (HSCs) differentiate into immature B lymphocytes that possess their own unique antigen receptor. Of these cells, those that yield specificity to antigens presented by endogenous cells are destroyed to prevent autoimmunity. Those who are incapable of spontaneously forming bonds with endogenous tissues further differentiate into mature, inactive lymphocytes. These cells circulate throughout the bloodstream until they encounter a foreign antigen that stimulates an intracellular signaling pathway to promote antibody production and cell division, creating many clones that can produce the same antibody [12].

Activated B lymphocytes can either mature into memory B cells, which exist in a quiescent state until they once again encounter their activating antigen (serving as a "memory" for the immune system), or plasma B cells which actively produce and secrete large volumes of antibody. The production of mAbs relies on the isolation and subsequent expansion of a single plasma B cell; unfortunately, primary B-cells cannot reach high cell densities and can only divide a finite number of times, making them incompatible candidates for swift and large-scale production of mAbs [13]. The revolutionary invention of hybridoma technology by Georges Kohler and Cesar Milstein in 1975 addressed this problem and paved the way for the industrial production of mAbs seen today [14].

2.3 Early Development: Hybridoma Technology

Hybridoma technology is a conceptually simple process of fusing plasma B cells with immortal myeloma cells, creating "hybridoma" cells. These cells possess the antibody-secreting capabilities of the plasma B cells while adopting the immortality and high proliferation of the cancer cells. This technique can be segmented into six main steps: first,

a mouse is injected with a specific antigen to invoke its adaptive immune response. This antigen is typically a key biomarker for a disease of interest that researchers are attempting to develop a targeted therapy for. After allowing infection to occur, spleen tissue is harvested from the mouse and the antibody-secreting B cells are subsequently isolated. Next, the harvested B cells must be fused with myeloma cells to generate hybridoma cells. Successfully fused cells must then be selected for by culturing in hypoxanthine-aminopterin-thymidine medium (HAT) medium.

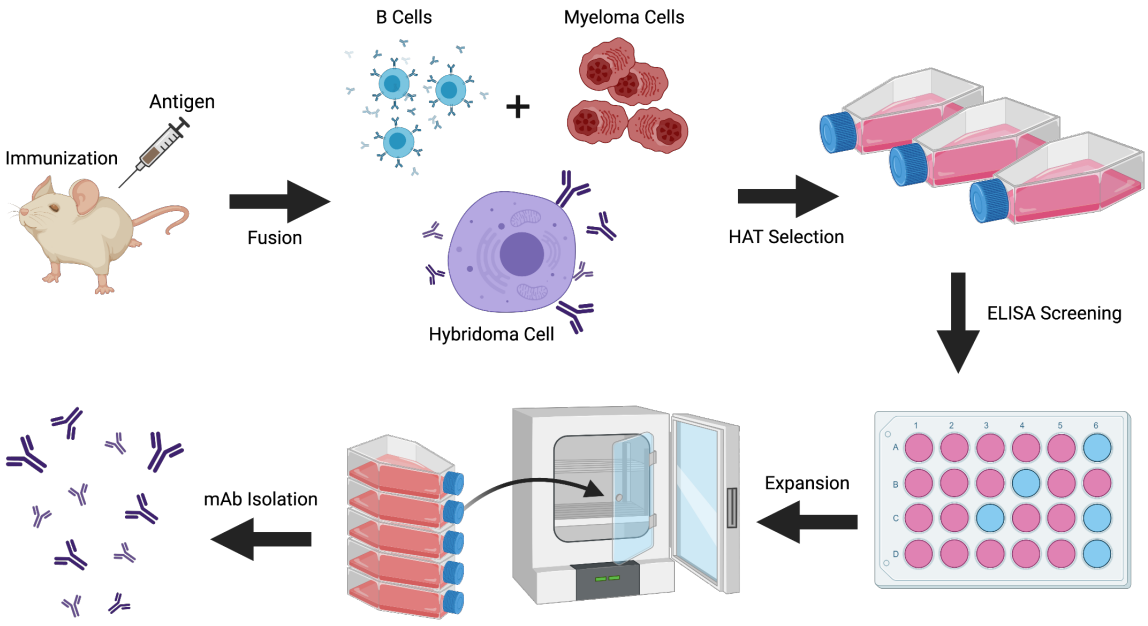


Figure 2.6: Hybridoma Technology (created using BioRender.com) [15]

This selection technique functions based on two key principles: unfused B cells from the spleen will not replicate in culture, and unfused myeloma cells are actively selected against by HAT in the medium. HAT blocks the DNA de novo nucleotide synthesis pathway, ultimately preventing DNA replication from proceeding; therefore, the only cells that will

survive are properly fused hybridoma cells. Next, hybridomas that selectively bind to the injected antigen are screened for using an antigen-presenting affinity assay (such as ELISA). Once the preferred clones are selected, they can be independently expanded in culture for the eventual harvest of their antibodies. All described steps are visualized in figure 2.6. The first FDA approved monoclonal antibody produced by this technology was Orthoclone OKT3 (generically known as muronomab-CD3), an immunosuppressant targeting the CD3 receptor developed by Janssen-Cilag in 1986 [16].

2.4 Principles of Monoclonal Antibody Production

Before mAbs can be manufactured at large, high affinity clones capable of neutralizing a specific disease epitope must first be isolated from B cells. This requires the sorting and screening of individual B cells from a population (such as spleen tissue or a PBMC patient sample), as described in the discussion of hybridoma technology. There are multiple existing techniques for clonal selection, with great potential for the development of novel methods and technologies. This is a critical step of the mAb manufacturing process because the efficacy of mAb therapeutics hinges significantly on the affinity of the antibody.

2.4.1 Clonal Selection

There are multiple laboratory technologies capable of high-throughput screening of populations of antibodies against a single antigen of interest, and most are based on assays that test the physical affinity between the two molecules. Phage display is an *in vitro* technology that involves the insertion of a protein-encoding gene into a bacteriophage, which then presents the protein on the surface of its coat, making it functionally similar to

MHC presenting an antigen [17]. These protein-presenting bacteriophages can then be screened against different proteins immobilized to a surface (such as a well plate). Spontaneous binding activity followed by subsequent washing allows for isolation of bound bacteriophages which can be sequenced to identify the gene that encodes for the high-affinity antibody, shown in Figure 2.7. This technique was pioneered by Sir Gregory Winter in the 1990's and has enabled the development of multiple FDA-approved humanized antibodies since the early 2000's; he ultimately received a Nobel Prize for this work in 2018 [18].

Another method of identifying high affinity therapeutic mAbs is to screen for them in human primary cells. This technique allows for the identification of fully human mAbs against a specific disease antigen by collecting a blood sample from a patient that has exhibited a successful recovery, negating the need for humanization [11]. First, B-cells (which make up approximately 5-10% of a PBMC sample) must be isolated using automated sorting techniques, such as flow cytometry [19]. The obtained population of B-cells is heterogeneous and therefore must be further sorted into single cells. The antibodies produced by these single cells can be screened for using techniques previously discussed. Fully human mAbs may also be screened for in transgenic mice hybridomas, which have been genetically modified to produce antibodies with the structural makeup and post-translational modifications of mAbs produced by humans. These techniques are also shown in Figure 2.7.

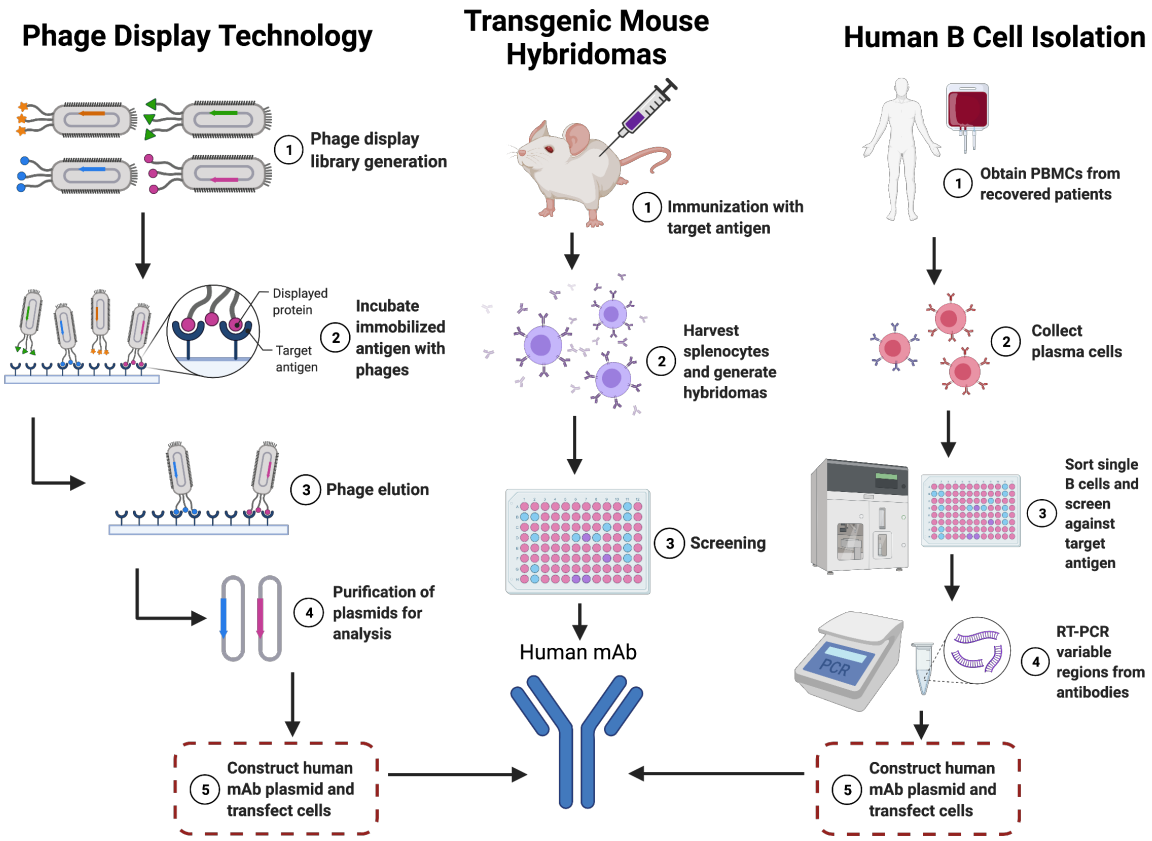


Figure 2.7: Human mAb Development Techniques (created using BioRender) [11], [20]

2.4.2 Creating an Expression Plasmid

After selecting the desirable mAb clone, the next step is to create robust mammalian cell factories that can produce large enough amounts to meet high industry demand. Immortal cell lines with well-characterized nutritional needs and metabolisms must be transfected with an expression plasmid for the mAb of interest. To achieve this, the gene that encodes for the variable regions of the light and heavy chains of the mAb must be amplified and sequenced; cDNA containing this sequence can be inserted into an expression plasmid alongside a promoter and a selection gene, which can be used to transfect robust mammalian cell lines for mAb production (Figure 2.8) [21].

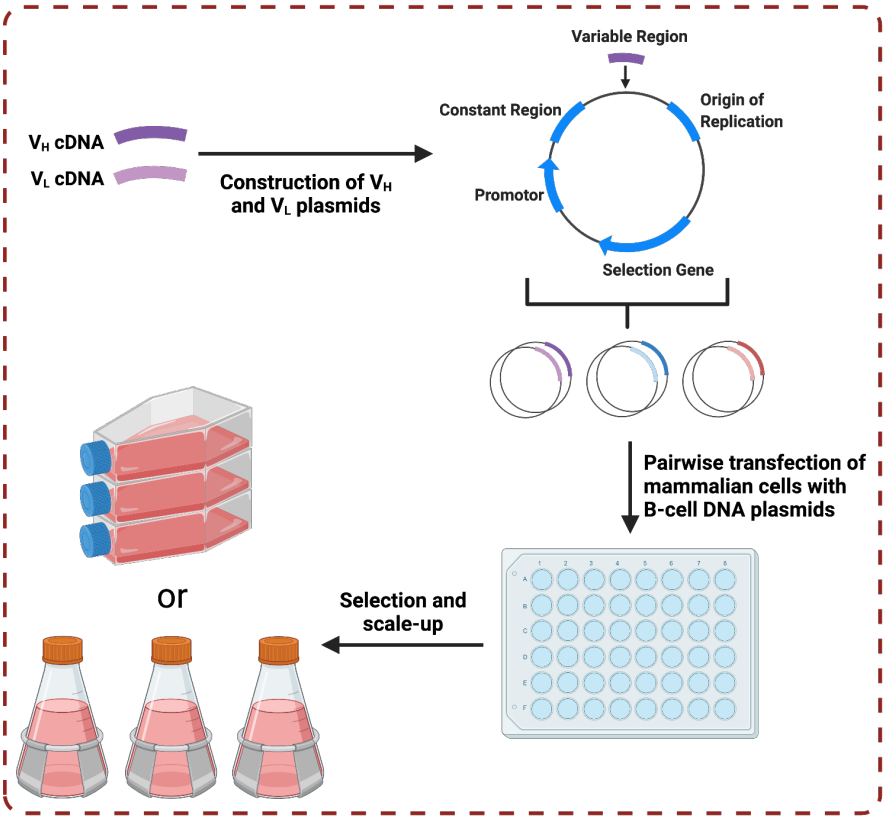


Figure 2.8: Creation of Human mAb Plasmid (created using BioRender.com)

2.4.3 CHO and NS0 Cells

The production of full-length therapeutic mAbs is largely carried out using two cell types – Chinese hamster ovary (CHO) and non-secreting murine myeloma (NS0) cells. CHO cells account for over two-thirds of industrial mAb production alone, while NS0 cells account for most that remain [22]. While both cells can exhibit stable growth in serum-free suspension cultures, CHO cells are comparatively more robust and can express mAbs that possess human-like posttranslational modifications. One downfall of NS0 cells is that they incorporate certain molecules into their glycosylation patterns that are immunogenic in humans, limiting their efficacy [23]. In comparison, CHO cells have been shown to incorporate lower levels of immunogenic molecules, rendering them more practical for the

production of biologic therapeutics for humans. Additionally, CHO cells are bereft of viral entry genes that make NS0 and other mammalian cells susceptible to viral infection, enhancing their robustness by mitigating this risk in cultures [24]. These advantages have led to the rapid characterization of CHO cells, including genomic sequencing in 2011 and the publication of extensive omics analysis in the years immediately following [25]. These advancements have crafted a formidable argument for the selection of CHO cells (and to a lesser extent, NS0 cells) for the synthesis of mAbs using engineered expression plasmids for high-affinity clones.

2.4.4 The Glutamine Synthetase (GS) Selection System

CHO and NS0 cells lack the natural ability to perform the catalysis of glutamine from ammonia and glutamate, so their culture media must be supplemented with it in order for them to survive [26]. This characteristic has been exploited for the specific selection of cells that are capable of mAb production by coupling a gene for the glutamine synthetase enzyme with the mAb gene in the expression plasmid [27]. Therefore, when cultured in glutamine-free media, only cells that have been successfully transfected will continue to survive and proliferate. Cells that are particularly high-producing can be further selected for by the incorporation of L-methionine sulfoximine (MSX) in the culture media, which is known to inhibit the activity of glutamine synthetase. The DHFR selection system is also widely used in the biopharmaceutical industry, but it is not relevant to this thesis because only GS-CHO and GS-NS0 cell cultures were analyzed [28].

2.4.5 Scale Up of mAb-Producing Cell Cultures

Mammalian cells that produce therapeutic mAbs are grown in suspension cultures and are capable of very high cell densities while maintaining above 90% viability. To accomplish this, a cell line that is capable of consistent, high production of a mAb of interest must be constructed. First, an industrial mammalian cell line with an existing selection system (ex: GS-CHO) is transfected with an expression plasmid designed for a specific mAb. The transfected cells are then cultured in selection media and expanded over the course of multiple weeks. The highest-producing clones are identified and selected for further expansion during this process [29]. Once the ideal clones have been iteratively selected for and expanded over weeks, small-volume aliquots are placed in cryopreservation tanks to create a master cell bank (Figure 28.3A). Individual vials are taken from this master cell bank for the creation of working cell banks. Cryopreserved cells are typically first inoculated in small-volume shake flasks in batch culture. Once a higher cell density in flasks has been achieved, the cells are passaged into a small seed bioreactor and cultured to high density once again. This process is iteratively repeated with larger and larger seed bioreactors until they are finally passaged into the final production bioreactor, which is typically hundreds or even thousands of liters in volume (Figure 28.3B) [30]. After the final culture falls below an undesirable viability threshold, cells and solid debris are removed via centrifugation and filtration, and the recovered spent media is preserved for downstream processing (namely mAb purification) [31].

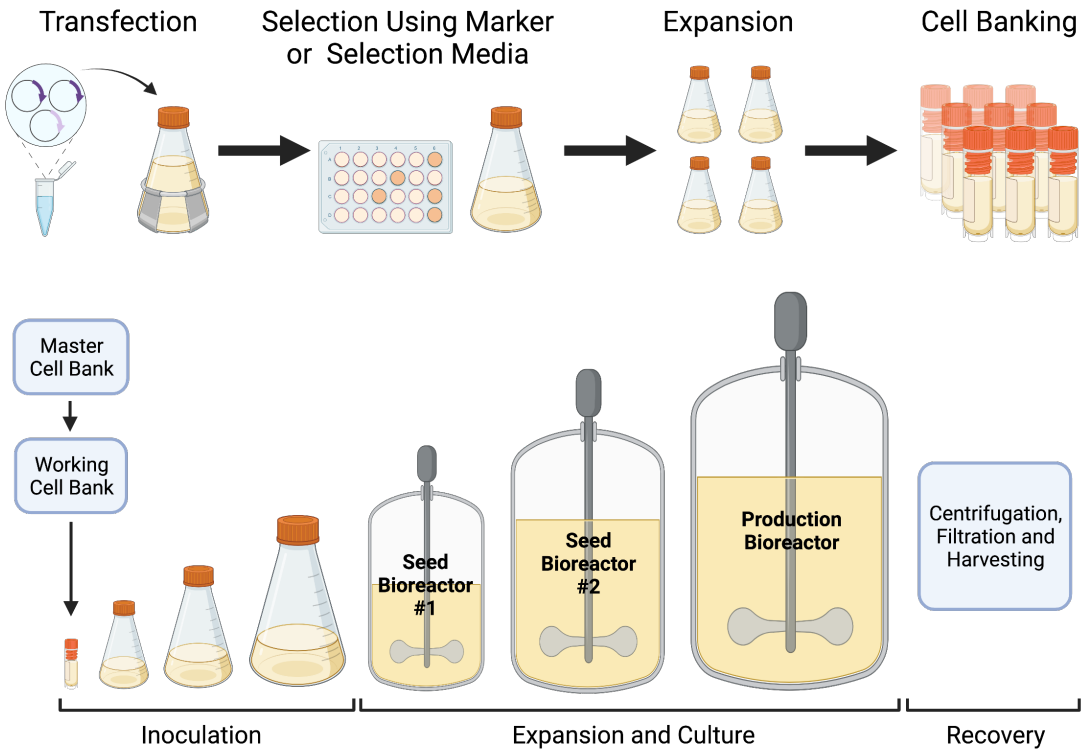


Figure 2.9: Clonal Selection and Scale-Up (created using Biorender.com) [30]

2.5 Current Manufacturing Practices and Future Optimization Strategies

Many improvements have already been implemented to enhance cell density and viability, as well as mAb titer and quality. For example, transitioning to serum-free media greatly improved the reproducibility of cultures by eliminating a large source of variability in media composition [32], and predictive mathematical modeling of cellular processes has been shown to save significant time and resources in the optimization of culture conditions [33], [34]. This section briefly summarizes broadly important advances that have been made and are continually being developed for industrial mAb production.

2.5.1 Single-Use Bioreactors and Improving Process Monitoring and Control

The implementation of bioreactors in place of shake flasks alone introduces significant

advantages in mAb-producing mammalian cell culture, but the size, build and functionality of bioreactors is very diverse. In a more recent development, many bioreactor manufacturing companies are transitioning their focus towards single-use bioreactors instead of reusable stainless steel. In a large-scale industrial setting, single-use bioreactors are attractive because they greatly reduce the risk of contamination [35]. With stainless steel bioreactors, the vessels must be extensively cleaned and sanitized between cultures to avoid contamination, and there remains risk. Single-use plastic bioreactors are sterilized before or on arrival and are disposed of after a single use, reducing the time and resources required for cleansing and sanitation as well as eliminating risk of contamination in the next culture. The employment of single-use sterile connectors for tubing and ports in bioreactors are also a necessary inclusion for the maintenance of sterility during inoculation, sampling and feeding of a culture.

Peripheral technologies for the real-time online monitoring and control of important process parameters such as pH, gas cascade intensity, dissolved O_2, temperature and agitation speed (for suspension cultures) make it increasingly easier to maintain robust and automated process control with the development of user-friendly and industry-standard software. The continuous improvement of bioreactor design and process control and monitoring technologies will allow for enhanced and reproducible production of mAb products.

2.5.2 Fed-Batch, Continuous and Perfusion vs. Batch Cultures

In an industrial setting, batch cultures are far and away the most widely employed for mAb production due to their simplicity and long history of success. However, fed-batch cultures

improve on them in principle by supplying key nutrients to prevent depletion, but this increases the culture volume. Continuous cultures employ a similar strategy, but an equal volume being added is removed beforehand to avoid increasing the working volume and diluting the culture; however, this results in a decrease in total viable cell number at the time of feeding. Perfusion culture techniques improve upon this by filtering out spent media and cell debris while retaining whole cells. A visual comparison of these culturing methodologies is shown in Figure 28.3C.

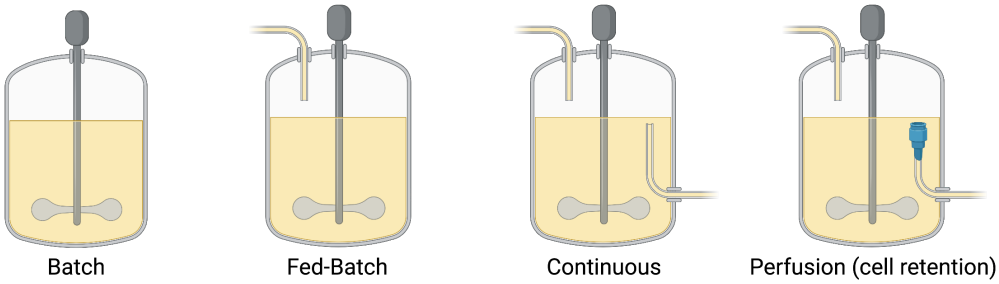


Figure 2.10: Modes of Culturing (created using BioRender.com) [36]

Implementation of these techniques is very sensible on paper, but actual incorporation into industrial production processes requires extensive development time and costs, as well as optimization – which is why batch cultures are still preferred by many businesses [1]. However, the long-term sustainability and prospects of fed-batch and continuous technologies have pushed some large biotechnology companies who are able to absorb the high upfront costs and technology transfer to transition more of their processes away from batch methodologies [1].

2.5.3 Omics Characterization

Comprehensive characterization of cell lines is a critical factor in the optimization of their culturing. Developing a greater understanding of the compositional and behavioral characteristics is a necessary first step in better accommodating their growth and productivity needs by means of optimized process parameters. Metabolomics broadly describes the large-scale quantification of small molecules that serve as substrates, intermediates, or products of metabolic processes in the cell – together, these make up the metabolome. Liquid chromatography-mass spectrometry (LC-MS) is the most widely applied method for metabolomics, but other technologies such as nuclear magnetic resonance (NMR) spectroscopy and gas chromatography-mass spectrometry (GC-MS) [37].

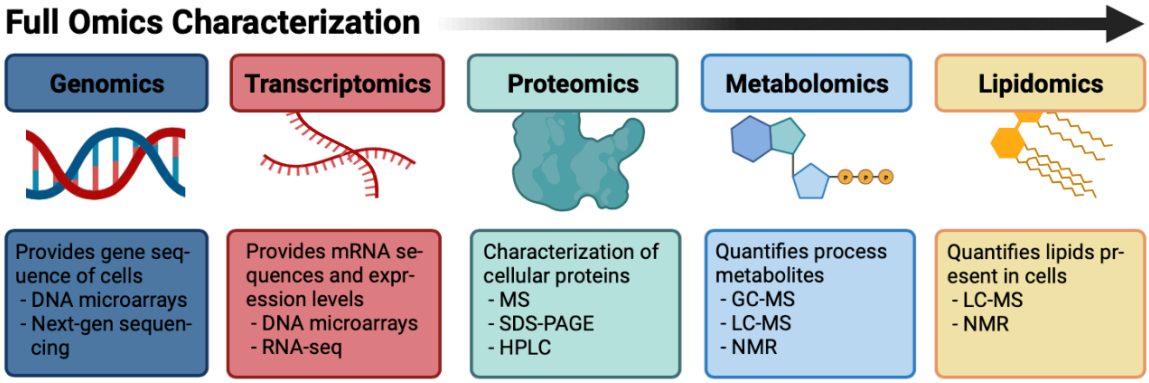


Figure 2.11: Omics Characterization of Cell Lines (created using BioRender.com) [38]

Proteomics similarly describes the large-scale identification and quantification of proteins that are produced by a culture of cells. Since proteins are expressed for a variety of purposes, proteomics can be used to characterize both the functional and structural proteomes. Gel electrophoresis and SDS-PAGE are commonly used for protein expression

profiling, and mass spectrometry used in tandem with chromatographic methods (such as liquid chromatography) is used for the isolation and structural characterization of proteins [39].

Genomics and transcriptomics, which entail the mapping and characterization of a cell's DNA and transcribed RNA through means of PCR, DNA microarrays, next-generation sequencing and RNA-sequencing can help further develop the understanding of a specific cell line's genome and functional transcriptome at key developmental and metabolic stages.

Lipidomics describes the chromatographic separation and subsequent identification and quantification of specific lipids by means of mass spectrometry or NMR. This study is important because, like proteins, lipids contribute significantly to not only structural composition (namely phospholipid membranes) but also cellular metabolism. All of these "omics" techniques, visualized in Figure 2.11, can be employed to develop comprehensive maps of cellular characteristics during different stages of both successful and failed cultures. Characterization on such a large scale requires collaborative effort between international research communities, of which projects like the regularly updated CHOgenome.org for the conglomeration of CHO genomic data are a prime example [40].

2.5.4 Media Design and Feeding Strategy Optimization

Culture media serves the purpose of providing sole exogenous support of key nutrients and macromolecules to cultured cells, and therefore its composition is extremely important to the optimization and success of mAb manufacturing. The relative concentrations of vitamins, amino acids, buffer salts, lipids and trace metals can yield significant impact on

the productivity and viability of cultured cells, and many cell lines are very sensitive to any compositional deviations. It has been found that the substitution of serum-supplemented media in favor of serum-free chemically defined media greatly enhanced the reproducibility of cultures by eliminating a significant source of compositional variation between bottles of the same culture media, which is especially critical for the maintenance of robust bioprocessing [32]. However, given that no two cell lines have identical nutritional needs, an important step in the optimization of mAb production is the determination of media composition that is tailored to meet the specific growth and productivity requirements of any given industrially relevant cell line.

This optimization can be informed by experimental observation of the dynamic consumption and production of media components and cellular metabolites in tandem with "omics" data described previously. The identification of significant shifts in a culture's metabolic behavior under different conditions elucidates either the excess or deficiency of key nutrients - both of which can result in growth inhibition, poor productivity, and the accumulation of toxic metabolites such as lactate and ammonia. It has been shown that media optimization alone can result in significantly enhanced volumetric productivity of mAb-producing cell lines by precisely meeting the nutrient requirements for cellular maintenance and proliferation, mAb production and energy metabolism [34].

2.5.5 Design of Experiments Methodology

Intracellular biochemical processes have been extensively mapped to better understand their needs and behavior, and researchers have also been able to apply mathematical equations to describe their dynamic relationships. This is one of the most important and

fastest-growing areas of bioprocess development and optimization because it enables *in silico* optimization through predictive modeling as opposed to empirical approaches like design of experiments (DOE) methodology. In short, DOE is an optimization strategy that uses applied statistics to generate an array of different experimental conditions that result in the comprehensive testing of multiple process parameters in a highly controlled manner. This systematically evaluates each parameter and its impact on the process responses (such as mAb yield or cell density) to maximize productivity and minimize variability. A technique known as factorial design is used to determine which factors (synonymous to independent variables) are to be varied during experimental runs and how many levels will be tested (ex: low, medium, high). Each experiment, or run, involves the application of a unique combination of factor levels, referred to as a treatment. Below is an example of a design matrix, which serves to visually depict the treatments to be applied for each run in an experiment. In this case, the authors employed 3^3 factorial design because three factors were tested (cell concentration at induction, $CuSO_4$ concentration at induction, and harvest time) each at three different levels, in an attempt to maximize percent menin expression in *Drosophila melanogaster* S2 cells [41]. It is important to note that while there are 27 possible treatments for this design, only 10 runs were carried out by the authors; this is an example of fractional factorial design, in which a subset of treatments from the full factorial design are chosen to emphasize importance on the variables most likely to have the greatest impact while minimizing time and costs. Screening experiments can be employed to identify significant factors to focus on for fractional factorial design, as is very commonly observed in clinical research [42].

Table 2.1: Example of Factorial Design

Culture	Cell concentration at induction (10^7 mL^{-1})	$CuSO_4$ concentration at induction (mM)	Harvest time days (post induction)	Menin (%)*
1	0.5	0.2	1	3.93
2	0.5	0.7	1	2.35
3	0.5	0.2	6	1.29
4	0.5	0.7	6	52.08
5	1.5	0.45	3.5	69.52
6	1.5	0.45	3.5	18.35
7	2.5	0.2	1	100
8	2.5	0.7	6	3.56
9	2.5	0.2	6	1.73
10	2.5	0.7	1	84.80

2.5.6 Model-Based Optimization

While DOE-based optimization is ultimately successful, it often requires that hundreds of potentially unsuccessful runs are conducted for comprehensive testing, which is costly in terms of both time and resources. Conversely, model-based optimization saves time and resources by mathematically simulating experimental results and outcomes based on known and derived equations that describe cellular behavior. These equations can reflect known mass balance principles for various biomolecules and their required proportional uptake for key cellular functions such as structural maintenance, mAb production and energy metabolism [34]. For example, the biomass composition of a specific cell line can be experimentally determined by the isolation and digestion of dry cell mass and sequential processing for quantification of amino acids, proteins, lipids, and carbohydrates content per cell. Similarly, the molecular composition of monoclonal IgG antibodies can be determined, and published information is already available. Applying these biomolecular

mass requirements in combination with known metabolic reactions allows for the dynamic predictive modeling of cellular allocation of nutrients, mAb productivity, proliferation, and toxic metabolite accumulation. Many different research groups and companies have recently made large strides within this area, and further experimental validation and evolution of computing capabilities will enable even greater potential for *in silico* bioprocess optimization. This will be critical in meeting the increasing demands of the expansive mAb market.

2.6 Multivariate Analysis Techniques

Since bioprocess optimization requires the consideration of many variables, multivariate analytical techniques are helpful in elucidating clusters of similarity between variables and between samples in a large dataset. Multivariate analysis is an umbrella term used to describe many statistical analysis methods for multivariate data, but the work in this thesis applies only two specific methods: principal component analysis (PCA) and hierarchical clustering analysis (HCA).

2.6.1 Principal Component Analysis

This analysis is a method for dimensionality reduction, meaning that is transforms a large dataset into a much smaller dataset of new variables that are orthogonally related and still capture a significant fraction of the variation in the original variables – these new variables are called principal components [43]. For this to be possible, the original variables must be standardized to be comparable, and this is performed by first calculating the z-score for each entry of data [44]. The equation for calculating the z-score is shown below:

$$z = \frac{value - mean}{standard\ deviation} \tag{2.1}$$

The z-score standardizes each variable by scaling them down to be comparable, giving each column of data a mean of zero and a standard deviation of 1. Due to the sensitivity of PCA, variables with much higher magnitudes would eclipse variables of smaller magnitude, making them appear much more significant than they likely are, so standardization is important in most cases to ensure the accuracy of PCA. Next, the covariance matrix *S* for the standardized variables must be computed using the following equation, where *n* denotes the number of samples included in the analysis [43]:

$$S = \frac{U^T U}{n - 1} \tag{2.2}$$

The eigendecomposition (extraction of eigenvalues and eigenvectors) of the covariance matrix *S* are then used to determine the principal components of the data. The eigenvalues are stored in matrix *D* and eigenvectors in diagonal matrix *P*, for the purpose of discussion. The values in matrix *P* are then sorted based on the magnitude of the corresponding eigenvector in matrix *D*, with the largest eigenvalue dictating the placement of its corresponding eigenvector in the first column of diagonal matrix *P* and iteratively repeating with the next largest eigenvalue until the matrix of eigenvectors is sorted in descending order [45]. Matrix *U* (which stores the original standardized variables) is then multiplied by re-ordered matrix *P* to produce the one-dimensional matrix of principal components, or PC scores; the values of the eigenvectors denote the PC loadings [45]. Geometrically, these principal

components are linear combinations of all of the original variables and are orthogonal from one another, meaning that each one is entirely independent. The first principal component is the one that is able to capture the most variation in the original data, with each following principal component capturing a diminishing amount.

An easy way to conceptualize PCA is to imagine that each variable is plotted on its own axis in space (such that ten variables would translate to ten theoretical dimensions). The first principal component is created by drawing a line through this ten-dimensional space that is as close to as many of the data points as possible, and the projections of the original data points onto this line represent their new values for this principal component. A new line is then drawn in an orthogonal direction from the first line that is, once again, as close to as many of the original data points as possible. This process is repeated as many times as desired. Each principal component captures a percentage of the total variance within the original data, and this percentage is very important when determining their significance. The power of PCA becomes clear when PCs are visualized together, and for the work in this project I opted to include only the first two, since these alone were able to capture a majority (>50%) of the variation in all cases. This visualization can elucidate clusters of similarity in the data, highlighting which cell culture samples are similar and different from one another when accounting for all variables, at the cost of explicit clarity regarding which original variables are the significant contributors. In this thesis, existing packages for RStudio were utilized to automate calculations and simplify the process of data visualization.

2.6.2 Hierarchical Clustering

Hierarchical clustering is an analysis that, similar to PCA, is used primarily to identify clusters of similarity in a standardized dataset. It can be performed using numerous methods which are broadly categorized into two groups: agglomerative (bottom-up) or divisive (top-down) [46]. Agglomerative algorithms initially consider every row or column of data points as an individual cluster, and iteratively form additional clusters based on a similarity metric. Euclidean distance was the selected metric for this work, and it describes the distance between two points in *n*-dimensional Euclidean space which can be calculated using the following equation, where *p* and *q* represent a row or column and *i* represents a single entry within each [47]:

$$d(p,q) = \sqrt{(p_i - q_i)^2 + \cdots + (p_n - q_n)^2} \tag{2.3}$$

The Euclidean distance between two rows or columns describes their similarity and is utilized to ultimately determine the linkage between clusters. Complete linkage criteria, selected for this work, requires that the distance between two clusters be defined by the maximum Euclidean distance that every pairwise combination of cluster elements can achieve. This means that the distance between clusters is determined not by their two closest elements (single-linkage) or the average of all elements (average-linkage), but rather their two farthest elements [48]. Combining the two most similar clusters is performed in a stepwise manner until all the data is merged into a single cluster that branches off into progressively smaller clusters, visually represented by a tree-like dendrogram. Complete linkage is one of the most commonly employed methods in hierarchical clustering algorithms and is

known to produce more compact clusters, which is helpful for this project in the endeavor to identify specific patterns and linkages in metabolic behavior [46]. Hierarchical clustering can be conducted horizontally or vertically across a data frame, which will be shown in the upcoming chapter. Dendrograms generated from the HCA are depicted along their respective axis against a heatmap of the standardized data matrix, which visualizes the time horizons of each variable using a color spectrum for easier pattern identification. All data must be standardized such that magnitudes do not impact the Euclidean distances.

2.7 Applications of Monoclonal Antibodies

Their high affinity and specificity yield great potential for many diverse applications of mAbs as a clinical tool. They have been shown to be capable of neutralizing a wide variety of diseases, but their usefulness extends beyond purely therapeutic applications. Monoclonal antibodies are also a powerful diagnostic tool and can even be used to augment existing medicines to enhance their efficacy by enabling specific targeting. A brief overview of the different application areas of mAbs will be discussed in this section, and a table of the most profitable mAb therapies is also provided.

2.7.1 Treatment and Diagnosis of Cancers

Oncology remains the most relevant and profitable sector of mAb applications with the highest recorded and projected sales of all other mAb application areas by far [49]. One proven mechanism is the targeting of human epidermal growth factor receptor 2 (HER2), an important component in breast cancer cell signaling, as shown by the commercially successful trastuzumab (developed and marketed as Herceptin by Genentech). Another

blockbuster Genentech therapeutic, Avastin (generically known as bevacizumab), targets vascular endothelial growth factor to similarly inhibit the cellular signaling of multiple cancers. Both drug products were within the top five grossing mAb biologics in 2018 [49].

Recent breakthroughs in mAb treatments for cancer immunotherapy by Merck and Co. and Bristol-Meyers Squibb selectively inhibit the programmed death 1 (PD-1) immune checkpoint, a surface protein on T and B cells which downregulates T cell inflammatory activity upon activation of the T cell receptor [50]. While PD-1 is important for homeostatic control of adaptive immunity, it is exploited by cancer cells to limit the immune response against them; thus, anti-PD-1 mAbs such as pembrolizumab and nivolumab are promising solutions [49]. Systemic administration of anticancer mAb therapeutics can be achieved by the exploitation of the enhanced permeability and retention (EPR) effect of tumors on their surrounding blood vessels, which essentially describes a leakiness which results from upregulated VEGF expression and weak structural integrity as a product of uncontrolled growth [51]. IgG antibodies are small enough to fit through the vessel gaps created by this phenomena, allowing for passive delivery of mAbs to the tumor microenvironment through systemic injection.

2.7.2 Monoclonal Antibodies as Immunomodulators

Immunomodulating monoclonal antibodies can be administered to treat other immune disorders outside of cancer. Drugs that selectively inhibit inflammatory cytokines such as tumor necrosis factor (TNF) have exhibited success in medicating against multiple autoimmune disorders, such as Humira (adalimumab) for the treatment of rheumatoid arthritis [52]. While Humira is by far the most commercially successful, there exist several

other TNF-blocking mAb drug products for the treatment of other autoimmune diseases such as ankylosing spondylitis, psoriasis and Crohn's disease [49]. The significance and prevalence of various interleukins, which are immunomodulating cytokines, have also made them an attractive target for autoimmune mAb therapies. Successful mAbs on the market which selectively target IL-12/IL-23, IL-17, IL-5 and IL-4a (to name a few) are used for the treatment of autoimmune diseases previously described in addition to others such as asthma and dermatitis [51].

Table 2.2: Top 10 Selling mAb Therapeutics

No.	Drug	Indication (First US FDA Approval Year)	Company
1	Adalimumab (Humira)	Rheumatoid arthritis (2002) Psoriatic arthritis (2005) Ankylosing spondylitis (2006) Juvenile Idiopathic Arthritis (2008) Psoriasis (2008) Crohn's disease (2010) Ulcerative colitis (2012) Hidradenitis suppurativa (2015) Uveitis (2018)	Abbvie
2	Nivolumab (Opdivo)	Melanoma (2015) Non-small cell lung cancer (2015) Renal cell carcinoma (2015) Head and neck squamous cell (2016)	Bristol-Myers Squibb
3	Pembrolizumab (Keytruda)	Melanoma (2014) Head and neck cancer (2016) Non-small cell lung cancer (2015) Lymphoma (2018) Cervical cancer (2018) Microsatellite instability-high cancer (2018)	Merck & Co.
4	Trastuzumab (Herceptin)	Breast cancer (1998) Gastric cancer (2010)	Roche
5	Bevacizumab (Avastin)	Colorectal cancer (2004) Non-small cell lung cancer (2006) Breast ERB2 negative cancer (2008) Renal cell carcinoma (2009) Glioblastoma (2011)	Roche

Table 2.2: Top 10 Selling mAb Therapeutics (Continued)

6	Rituximab (Rituxan)	Non-Hodgkin's lymphoma (1997) Chronic lymphocytic leukemia (2010) Rheumatoid arthritis (2006) Pemphigus vulgaris (2018)	Roche
7	Infliximab (Remicade)	Crohn's Disease (1998) Rheumatoid arthritis (1999) Ankylosing spondylitis (2004) Ulcerative colitis (2005) Psoriatic arthritis (2005) Psoriasis (2006)	Johnson & Johnson
8	Ustekinumab (Stelara)	Psoriasis (2009) Psoriatic arthritis (2013) Crohn's Disease (2016)	Johnson & Johnson
9	Eculizumab (Soliris)	Paroxysmal nocturnal hemoglobinuria (2007) Atypical hemolytic uremic syndrome (2011) Generalized myasthenia gravis (2017) Neuromyelitis optica spectrum disorder (2019)	Alexion
10	Omalizumab (Xolair)	Asthma (2003) Chronic idiopathic urticaria (2014)	Roche

2.7.3 Other Clinical Applications

While inhibitory monoclonal antibodies can exhibit therapeutic effects on their own, not all are neutralizing agents. Instead, they can be incorporated into drug product formulations to provide selective delivery for active pharmaceutical ingredients which are cytotoxic, which enhances efficacy and mitigates unwanted side effects. They are becoming increasingly relevant in nanoparticle-based delivery systems, because the encasing of the drug within the nanoparticle (which are often synthesized from organic polymers, lipids, or gold) offers enhanced stability during systemic circulation while the surface mAbs selectively bind to the disease antigen, offering and maintaining close proximity of the

cytotoxic drug to its target [53]. There is much work still to be done within this research area, but there are multiple nanoparticle-based drug-antibody conjugates in development that have shown promise in the treatment of breast and ovarian cancers [54] [55].

2.7.4 Diagnostic and Non-Clinical Applications

Monoclonal antibodies are an exciting addition to the pharmaceutical world, but they are also present in many diagnostic applications and common lab technologies. They serve as the backbone for enzyme-linked immunosorbent assays (ELISAs), which utilize mAbs that are conjugated to an immobilized antigen and coupled with a reporter enzyme that can detect the presence of a specific substrate, such as a hormone or antibody, when incubated with a biological sample [56]. ELISA is the technology most commonly incorporated in pregnancy tests but can also be used in a wide variety of other diagnoses. Monoclonal antibodies can also be found in Western blots, flow cytometry, fluorescent microscopy, and affinity chromatography columns [57].

CHAPTER 3

DEVELOPING AN ANALYTICAL PLATFORM USING EXISTING DATA

3.1 Introduction

Before experimental work for this thesis began, the multivariate analysis techniques were developed by practicing with existing data collected by a postdoctoral researcher in our lab. Quiroga *et al.* cultured GS-NS0 cells for the development and validation of a predictive mathematical model to optimize culture media and feeding strategies for fed-batch cultures of mammalian cells, based on over one hundred metabolic, ATP and mass balance equations [34]. The model is reliant on data collected from an initial batch culture, after which it performs parameter estimation to output a recommendation for basal media supplementation and a feeding strategy for a fed-batch culture of the same cell clone. These conditions were then experimentally applied in an "unoptimized" fed-batch culture, during which more data is collected for the model. A new media supplementation and feeding strategy was output, which were applied in an "optimized" fed-batch culture to enhance its longevity and, ultimately, mAb titer. Quiroga *et al.* reported in their 2018 paper that model-based optimization successfully increased the mAb titer in a fed-batch culture of GS-NS0 cells by as much as 167% and extended the culture longevity by two days [58]. The data from the unoptimized and optimized cultures will be compared using unsupervised multivariate analysis techniques.

3.2 Aim and Objectives

The aim of this chapter is to successfully program a multivariate analysis platform which

is capable of receiving experimental data and automatically computing and visualizing results from HCA and PCA. To achieve this, a software tool must first be selected that is capable of preparing the data and performing the required analysis. Specifically, a method for data standardization by calculating z-scores from raw data will need to be programmed, and data split across multiple files (as was the case for the unoptimized and optimized cultures) must be joined. Then, an algorithm for calculating the principal components matrix will be programmed and the percent of variance captured by each will be subsequently calculated. An agglomerative hierarchical clustering method using a Euclidean distance similarity metric and complete-linkage criteria must also be programmed to generate both horizontal and vertical dendrograms for the standardized data matrix. Lastly, a method or package will need to be developed or selected to clearly visualize the results from the PCA and HCA. The data collected by Quiroga *et al.* will be analyzed using this developed platform, which will hopefully validate the findings reported in their publication as well as potentially extract new findings.

3.3 Materials and Methods

3.3.1 Programming in RStudio

RStudio software was chosen due to the lack of a paywall and extensive availability of existing resources for learning and programming. RStudio is the most widely used software for writing and running code using the R programming language, which was designed for statistical analysis and data visualization.

3.3.2 Data Input and Processing

The data was exported from Excel into .CSV format in order to be input and processed in RStudio using the read.csv() function, which places the rows and columns into a dataframe as they were arranged in the Excel file. Two separate data frames were created - one each for the unoptimized and optimized culture data. Each set of data represents the mean values calculated from three replicate cultures and measured 26 independent variables: all 20 amino acids sans glutamine (mM), ATP (mM), viable/apoptotic/dead cell densities (cells/mL), glucose (mM), lactate (mM) and mAb titer (mg/L). All values that were zero were replaced by $1x10^{-20}$, being on average 17-20 orders of magnitude smaller than the smallest nonzero value in the column to avoid problems with dividing by zero during standardization. Then, datasets were merged and standardized using the unite() and scale() functions respectively. The scale() function standardizes the data using the z-scoring methodology previously described in section 2.6.1, and unite() appends one dataframe to the other. There were 28 observations of each independent variable per unoptimized culture and 24 observations recorded for each optimized culture – therefore, in the final averaged and merged dataset analyzed in this chapter, there were a total of 52 observations for each of the 26 independent variables.

3.3.3 Hierarchical Clustering Analysis

Hierarchical clustering of the data was performed by nesting the dist() function, which calculates the distances between each row of data, inside the hclust() function to generate a dendrogram that sorts the samples based on their similarity across all variables. A condition was provided inside the dist() function to specify Euclidean distance as the desired

metric. The generated dendrogram was then displayed alongside a heatmap that was created using the heatmap.2() function in the "ggplots" package. A function argument specifying the inclusion of a dendrogram along the vertical axis, which sorts variables based on the similarity of their z-score profiles to one another, was also added. The z-score for each data point is visualized with a red/blue gradient in the heatmap such that dark blue represents a z-score that is significantly less than the column's mean and dark red represents a z-score significantly greater than the column's mean. The largest and most significantly different clusters have been boxed and color-coded for ease of visualization.

3.3.4 Principal Component Analysis

The principal components matrix was generated using prcomp() from the "stats" package. The percentage for each principal component was calculated using equation 3.1, where *i* denotes either the first or second PC:

$$PC_i\ \% = \frac{(Standard\ Deviation\ of\ PC_i)^2}{\sum_{i=1}^{i=2} PC_i^2} \times 100 \qquad (3.1)$$

PCA visualization was performed using a scatterplot, with PC1 on the x-axis and PC2 on the y-axis. To generate the graphs, the geom_point() function in the "ggplot" package was used, and the result is displayed in Figure 3.3.

3.4 Results

3.4.1 Hierarchical Clustering Analysis

Figure 3.1 visualizes the results of the hierarchical clustering. The color-coded boxes correspond with the boxes in the plots for viable cell density in Figure 3.2. The horizontal labels are sample IDs which denote whether the sample was taken from the unoptimized or optimized culture, followed by the number of hours after the start of the culture that the sample was taken. The vertical labels indicate the measured variables that correspond with each column.

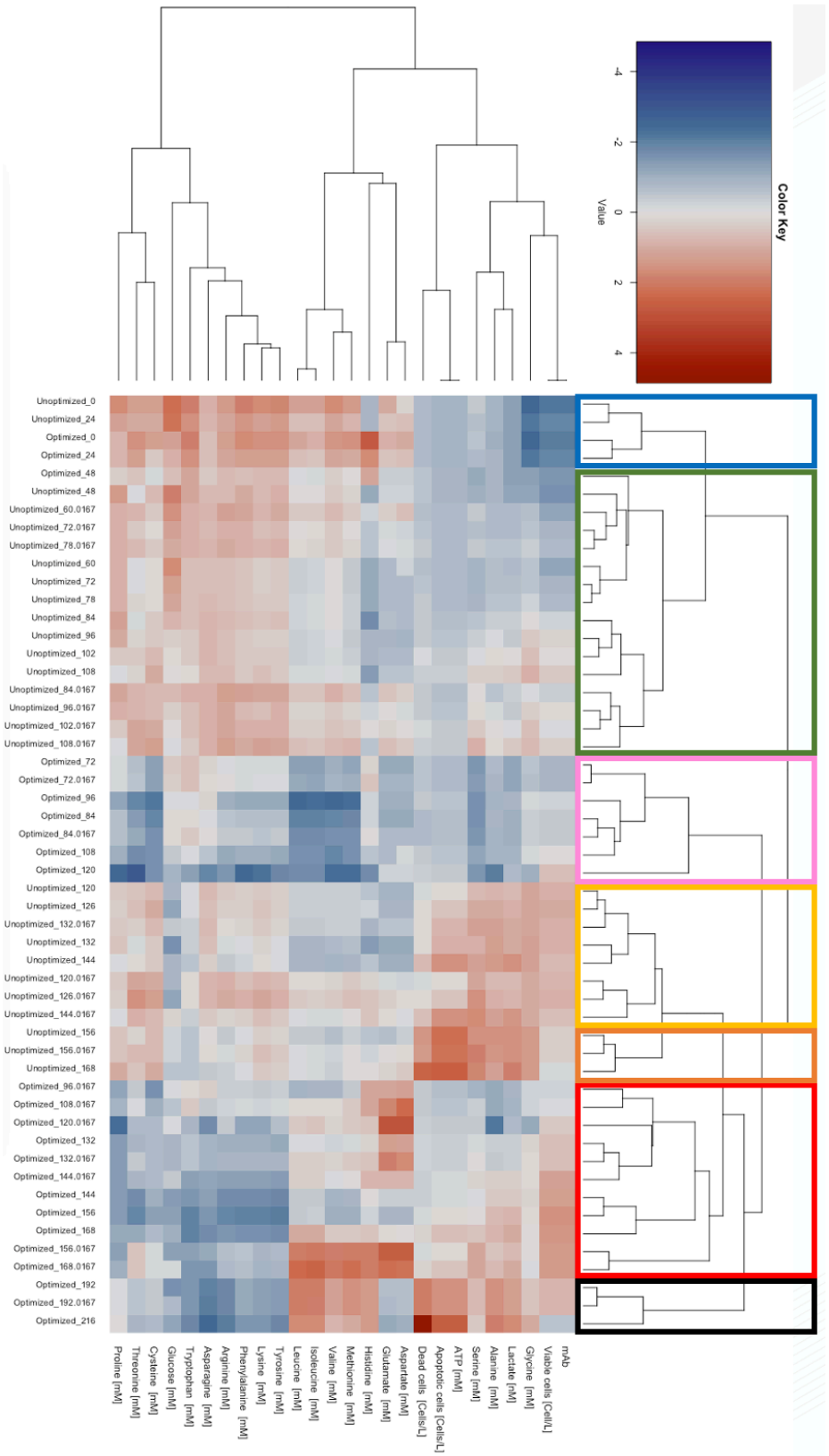


Figure 3.1: HCA of Unoptimized vs. Optimized NS0 Fed-Batch Cultures

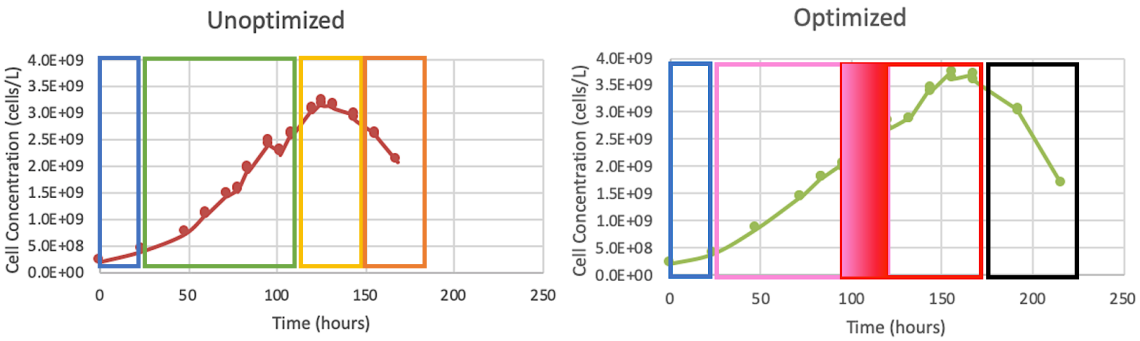


Figure 3.2: Viable Cells for Unoptimized vs. Optimized NS0 Fed-Batch Cultures

Initially, it must be noted that the major clusters of samples identified by the HCA correspond very closely with the four stages of growth commonly observed for cultured cells: lag, exponential growth, stationary, and decline/death. It can be seen that the lag phases of both cultures (first 24 hours) cluster together, but after 48 hours the concentration profiles of the two cultures diverged and were no longer statistically similar. Another notable observation is that the clusters for the unoptimized culture were exclusively sequential, meaning that there was no overlapping of clustered data points over time. This contrasts with the optimized culture, which as can be seen in Figure 3.2, possesses overlapping early and late exponential phases. Specifically, the feedings at 96 and 108 hours caused the next days' samples to cluster broadly with the samples collected during the early exponential phase (in pink) instead of the late exponential phase (in red), signifying that the optimized supplementation in these feedings reverted the cultural conditions to be more similar to what is observed during earlier stages of exponential growth. However, it is important to note that the early exponential phase for the optimized culture is metabolically distinct from that of the unoptimized culture and instead clusters with the late-stage exponential growth of the unoptimized culture, rather than its early-stage growth. Overall, the heatmap

displays relatively lower concentrations of many amino acids in the optimized media compared to their respective time points in the unoptimized culture.

3.4.2: Principal Component Analysis

The PCA results are visualized in Figure 3.3 on the following page. The clusters of data points have been manually color-coded and circled to correspond with the same clusters identified in the HCA. These results clearly show that there is a significant difference between the unoptimized and optimized fed-batch cultures across all of the collected data. First, it is notable that PC1 and PC2 collectively account for approximately 68% of the variation within the original dataset of 24 variables, bolstering the significance of the clusters in the graph. The key for the data points shown below the graph is ordered by culture, and each point is color-coded to correspond with the significant clusters identified in the HCA. This allows for clear visualization of the transition into the different growth phases within each culture and successfully showcases the shifting metabolic profiles between 96 and 127 hours in the optimized culture. It can be seen that after 48 hours, all of the data points for the optimized culture are significantly segregated from the unoptimized culture clusters. Interestingly, the most divergent clusters between the cultures are those representing the early exponential growth (denoted by the pink and green data points), highlighting the immediate difference between feeding strategies.

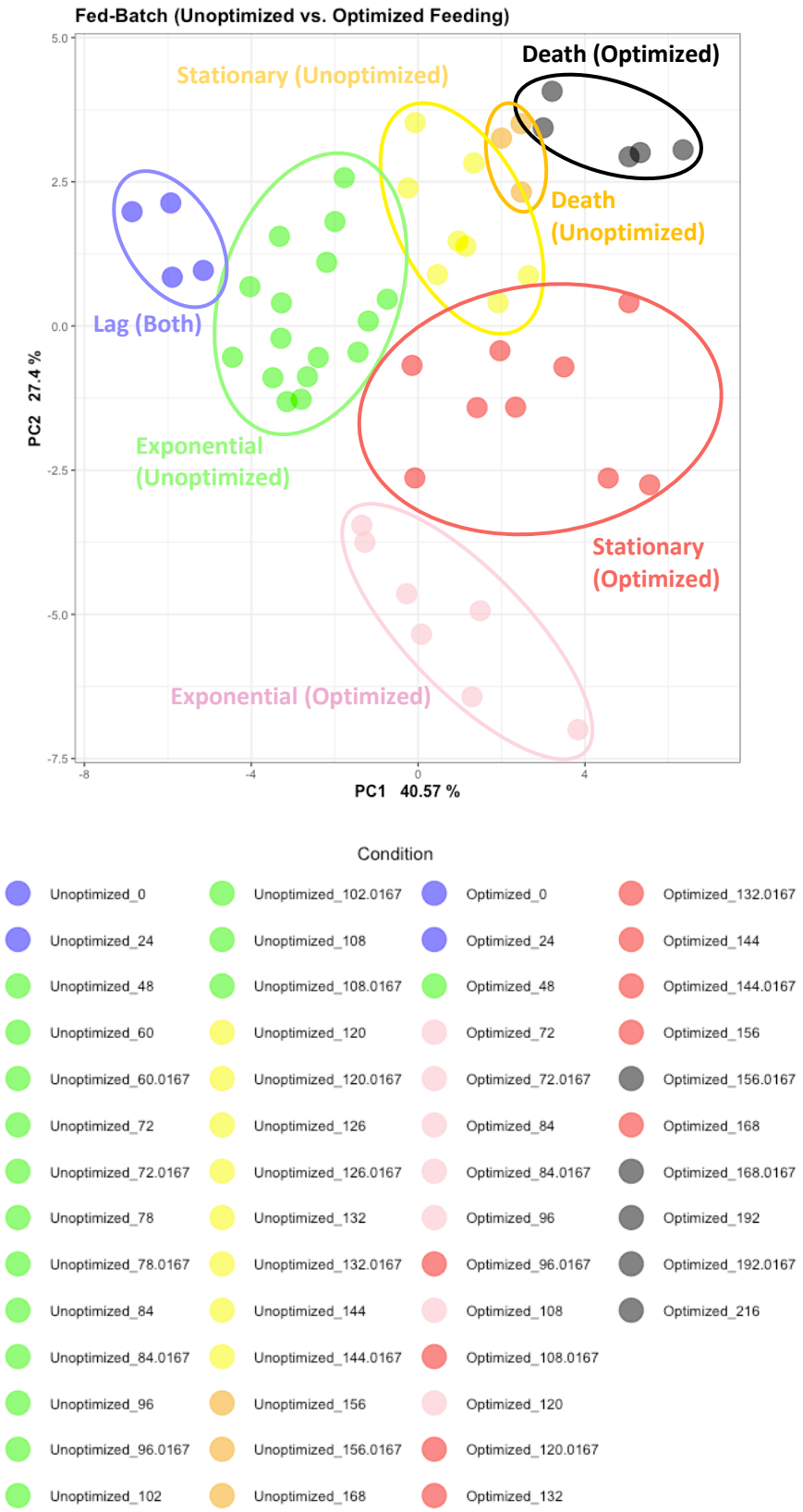


Figure 3.3: PCA of Unoptimized vs. Optimized NS0 Fed-Batch Cultures

3.4.3 Effect of Optimization on Culture Longevity, Density, and Productivity

The resulting viable cell densities and mAb titer for each culture are visualized in Figure 3.4. The optimized fed-batch culture exhibited increased longevity and maximum viable cell density, both clear benefits of the model-based media design and optimized feeding strategy. Enhanced benefits are also reflected in the mAb titer as well, as the optimized fed-batch culture achieved a maximum mAb concentration of over 300 mg/L while the unoptimized fed-batch culture achieved approximately 225 mg/L.

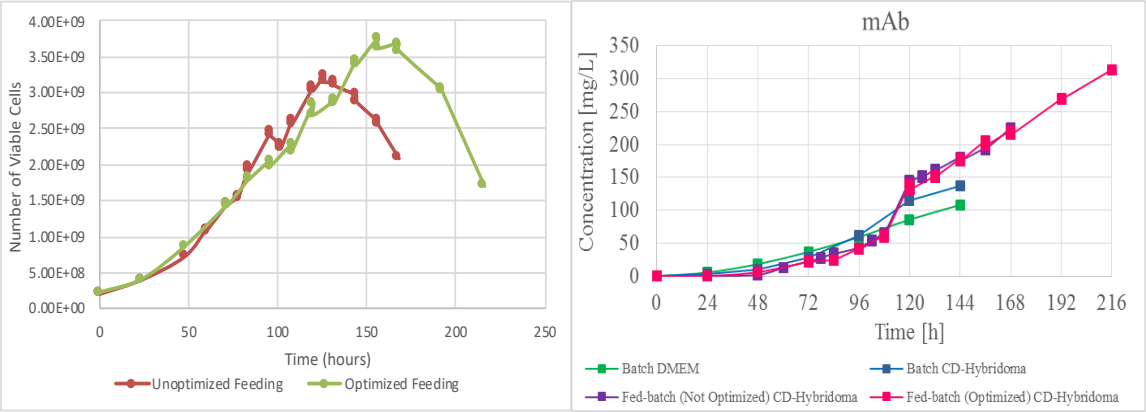


Figure 3.4: Enhanced Longevity and Productivity in Optimized NS0 Culture [34]

3.5 Discussion

The clustering of the lag phases in both the HCA and PCA is not surprising, given that neither culture has been supplemented with feedings yet and both cultures would need to similarly adjust to inoculation in the bioreactor under the same experimental conditions (sans basal media supplementation). Likewise, the divergence of the cultures not long after this point is also sensible given the timing that the feeding strategies were implemented. The overlapping of data points during the early and late exponential growth phases in the

optimized culture is significant because it shows that the optimized feeding strategy more significantly affected the metabolic behavior of the cells compared to the unoptimized strategy. This period of shifting between early and late exponential growth metabolic behavior is ideal because it shows that the culture is resistant to a permanent shift towards stationary growth, and earlier stages of growth are often when cells are the most productive because there are adequate exogenous nutrients to support rapid division and mAb production. The optimized feeding strategy was designed to limit the excess of nutrients in culture media, forcing the cells to sustain a more controlled metabolic state in moderate concentrations of nutrients. The advantageous results of the optimized media composition and feeding strategy compared to those that were unoptimized verifies existing observations that an excess of nutrients may initially promote rapid growth that is unsustainable and will ultimately induce a state of metabolic stress, negatively impacting culture longevity and volumetric productivity.

The ability to elucidate clusters of metabolic similarity within and between cultures from unsupervised analysis, which happen to broadly correspond to the phases of cell growth, is very interesting. This unbiased approach to the critical analysis of cell culture for the production of a specific mAb will be explored in greater depth in the coming chapters. GS-CHO cells will be cultured for the purposes of characterizing their metabolic behavior in chemically defined media. The identification of metabolic shifts will allow for the determination of key nutrients which could be supplemented into the basal media to extend culture longevity by optimizing cell viability and volumetric productivity.

By limiting the excess of nutrients during exponential growth, the optimized feeding strategy promoted differential metabolic behavior that reduced cellular stresses and

enhanced culture longevity by two days. This ultimately resulted in approximately a 40% increase in volumetric productivity, as shown in Figure 3.4. These benefits are a testament to the success of the model-based optimization developed and performed by Quiroga *et al.,* and the multivariate analytical platform applied to the data discussed in this chapter will be applied to new GS-CHO cultures moving forward to ultimately inform model-based parameter estimation to increase culture longevity, maximum viable cell density and mAb productivity.

CHAPTER 4

BATCH CULTURING GS-CHO CELLS FOR SPENT MEDIA ANALYSIS

4.1 Introduction

After programming PCA and HCA methods for cell culture analysis based on existing data, the next step was to apply these analytical methods for the optimization of new cultures. GS-CHO cells were selected for the following experiments because there was an available allotment in cryopreservation from a previous lab member's research. Additionally, GS-CHO cells have significant industrial relevance for the production of mAbs and other recombinant proteins. The specific GS-CHO clone utilized in this experiment produces a rituximab biosimilar which, as discussed in section 2.6.2, was the sixth highest grossing mAb therapeutic in 2018 [49]. The cultures were maintained for a total of 10 days because it was at this point that the viability for all flasks fell below 50%.

4.2 Aims and Objectives

The aim of this chapter is to present the methods for the GS-CHO batch cultures and discuss the specific analytical techniques utilized for the collection of multivariate data. Starting at the time of culture inoculation, spent media samples were collected, processed, and stored every 24 hours. The spent media samples collected and stored in the freezer daily were analyzed after termination of the culture flasks on day 10. The analytical methods described in the following pages serve the purpose of profiling the dynamic metabolic consumption and production patterns of the GS-CHO cells. Specifically, the quantification of amino acids, glucose, toxic metabolites (lactate and ammonia), vitamins, ATP and mAb

concentrations in spent media were conducted in tandem with daily collection of growth and viability data. The results for viable cell density, ATP availability per viable cell, mAb productivity, and lactate, glucose and ammonia concentrations will be shown and discussed in this chapter while the specific results for amino acids and vitamins concentration will be woven into the discussion in Chapter 5.

4.3 Materials and Methods

4.3.1 Creation of a Cell Bank

Multiple vials from the existing GS-CHO bank created by a previous student were thawed and cultured in small volumes, and unfortunately exhibited very low initial viability. Try-pan Blue assays performed immediately after multiple inoculations confirmed that initial viability was, on average, well below 50%. However, viability over 50% was eventually achieved after multiple media exchanges (performed every other day), and once a viability over 90% was achieved, the cells were expanded from a working volume of 20 mL to 200 mL. These cells were then transferred to cryopreservation media and banked at a concentration of 1.5×10^7 cells/mL in 1 mL vials for usage in experiments.

4.3.2 Inoculation and Culture Conditions

Culture media was prepared beforehand by supplementing 980 mL of CD-CHO media with 10 mL aliquots each of MSX and penicillin/streptomycin. For all experiments, two vials were briefly thawed at 37°C and inoculated in 20 mL of media in 125 mL Erlenmeyer flasks, resulting in a seeding concentration of 7.5×10^5 cells per mL. The cells were then passaged every two days into progressively larger volumes, first 60 mL and lastly in 200

mL with a seeding density of 4 x 10^5 cells/mL for every volume. Two days after passaging into 200 mL, each flask was split into two new 200 mL flasks to result in four total 1 L Erlenmeyer flasks containing 200 mL cultures at an initial density of 4 x 10^5 cells/mL. This seeding density was chosen because it was shown by a previous student in our research group to result in the highest maximum cell density during a 10-day culture when compared to lesser seeding densities [23]. These four flasks were kept in an incubator at 37°C, 5% CO_2 and horizontal orbital shaking at 120 RPM.

4.3.4 Daily Sample Collection

The cultures were maintained for ten days each, and a strict daily sampling routine was maintained. First, 500 µL were collected from each flask for the purposes of immediately assessing cell density, viability, and ATP concentration. Then, 2 mL was collected from each flask and centrifuged at 200g for 5 minutes. 500 µL was collected from the supernatant and transferred into a microcentrifuge tube with 10 kDa filter insert and centrifuged for one hour at 1000 g and 4°C. This filtration step is critical to remove any cells or larger molecules because the samples will be analyzed using high-performance liquid chromatography (HPLC) for the quantification of amino acids and vitamins. Another 500 µL sample was collected from the supernatant and filtered through a 45 µm filter attached to a syringe to remove any cells or cell debris for the purposes of HPLC quantification of mAb titer. Lastly, an unfiltered 700 µL sample of the supernatant was collected for the purpose of profiling metabolites using a benchtop BioProfile Flex 2 from Nova Biomedical. The flasks were then returned to the incubator and samples for HPLC and BioProfile analysis were stored in a -20°C freezer.

4.3.5 Determination of Cell Density, Viability and ATP Concentration

The fresh, unfiltered 500 μL samples not stored in the freezer were used for immediate assay of viable cell number and ATP concentration. Cell viability was measured using a benchtop Guava EasyCyte Flow Cytometer with Guava Viacount reagent from Luminex. The Viacount reagent employs one "nuclear dye", which stains nucleated cells (to distinguish cell debris from whole cells) and one "viability dye" (which distinguishes living and apoptotic cells from dead cells due to their differences in permeability). Samples were assayed in duplicates in a 96 well plate. Viability and apoptosis gates were consistent for all samples, and were manually adjusted at ten days, after the final culture samples were run, to ensure accuracy in capturing the apoptotic and dead populations of cells. In parallel, ATP quantification was conducted using benchtop Glomax Luminometer and a CellTiter-Glo Assay kit from Promega. The sample was mixed and incubated with CellTiter-Glo reagent and buffer in a 96 well plate and luminescence was subsequently measured. ATP concentration was determined from the luminescence readings using ATP standards to generate a calibration curve.

4.3.6 HPLC Quantification of mAb Production

The mAb concentration in supernatant samples was determined using HPLC. The machine used was an Ultimate 3000 system from Thermo Fisher Scientific, with the diode array detector (DAD) attachment for detection of amino acids using UV-Vis spectrophotometry. Machine programming and operation were conducted using Chromeleon 7 software. The MAbPac Protein A column from Thermo Fisher was used for the isolation of mAb product, and the programming of the instrument method was guided by the protocols detailed in

the product manual [59]. Mobile Phase A was prepared with 50 mM sodium phosphate and 150 nM of sodium chloride dissolved in one liter of MilliQ water and pH-adjusted to 7.5 with concentrated HCl. Mobile Phase B was prepared the same way, but a greater amount of HCl was added to adjust the pH to 2.5. A solvent wash of 90% Mobile Phase A and 10% acetonitrile used, and an additional solution of Mobile Phase A with 0.1% sodium azide was prepared for column storage. A constant flow rate of 0.5 mL/min at a detection wavelength of 218 nm was employed for 8 minutes per sample, with a 3 minute equilibration period in between samples. Serial dilutions of a Protein A standard were prepared for the generation of a calibration curve, and samples were run immediately thereafter. Antibody concentrations were calculated by integrating the chromatogram peaks and mathematically relating them to the calibration data. Specific productivity was calculated by subtracting the difference in concentration between each consecutive pair of samples and dividing by the number of viable cells.

4.3.7 HPLC Quantification of Amino Acids in Spent Media

To quantify the concentration of each of the 20 amino acids found in mammalian cells, an HPLC protocol was developed based on methods published by Agilent Technologies, the manufacturer for the Zorbax C18 column that was used [60]. Two mobile phases were prepared: the first was a mixture of 10 mM sodium phosphate, 10 mM sodium tetraborate, and 0.5 mM sodium azide pH-adjusted to 8.2, and the second was a mixture of 45% acetonitrile, 45% methanol, and 10% MilliQ water by volume. A constant flow rate of 1.5 mL/min was chosen because it minimized sample run time while maintaining clear separation of peaks. Two UV-Vis wavelengths were measured simultaneously during analysis:

338 nm (which was able to capture every amino acid except for proline), and 262 (optimized for the maximum absorption of proline). Calibration standards were prepared by supplementing a purchased standard mixture of pure amino acids with any that were not originally included by the retailer. Dilutions of these standards were run not only for the generation of a calibration curve but also for the identification of retention times for each amino acid. The methods published by Agilent provided a point of reference for me to conclude which peaks corresponded each amino acid, which allowed me to define a peak window within Chromeleon for each one; this essentially communicates the expected retention time for every amino acid (with some room for error) such that it can automatically perform peak identification for each one of the 40 samples run. Identified peaks were integrated by Chromeleon, and the calibration data was used to determine the concentration of each respective amino acid per sample.

Table 4.1: Amino Acid Specifications for HPLC Analysis

Peak Window #	Amino Acid Name	Retention Time (min)	Max Absorbance Wavelength (nm)
1	Aspartate (Asp)	1.740	338
2	Glutamate (Glu)	2.797	338
3	Asparagine (Asn)	5.040	338
4	Serine (Ser)	5.357	338
5	Glutamine (Gln)	6.100	338
6	Histidine (His)	6.353	338
7	Glycine (Gly)	6.703	338
8	Threonine (Thr)	6.920	338
9	Arginine (Arg)	8.037	338
10	Alanine (Ala)	8.300	338
11	Tyrosine (Tyr)	9.690	338
12	Cysteine (Cys)	10.943	338
13	Valine (Val)	11.757	338
14	Methionine (Met)	12.017	338

Table 4.1: Amino Acid Specifications for HPLC Analysis (Continued)

15	Tryptophan (Trp)	12.957	338
16	Phenylalanine (Phe)	13.390	338
17	Isoleucine (Ile)	13.603	338
18	Leucine (Leu)	14.303	338
19	Lysine (Lys)	14.713	338
20	Proline (Pro)	18.998	262

4.3.8 HPLC Quantification of Vitamins in Spent Media

The method utilized for vitamins quantification was designed around one that was published by Thermo Fisher Scientific for the simultaneous quantification of water and fat soluble vitamins [61]. For the purposes of this research, however, only eleven water-soluble vitamins were investigated – their retention times and maximum detection UV-Vis wavelengths are displayed in Table 4.2. Using the Thermo Fisher protocol as guidance, peak identification was carried out by running pure samples of each individual vitamin to determine the retention times relative to one another, as well as their optimal detection wavelengths. Serial dilutions of a mixture of all eleven vitamins were also prepared to obtain calibration data. Despite there being a substantial amount of heterogeneity in the optimal detection wavelengths for the vitamins communicated by Thermo Fisher, it was empirically determined here that just three wavelengths were sufficient to capture the peaks.

Table 4.2: Vitamin Specifications for HPLC Analysis

Vitamin Name	Retention Time (min)	Max Absorbance Wavelength (nm)
Thiamine (B_1)	1.806	260

Table 4.2: Vitamin Specifications for HPLC Analysis (Continued)

Ascorbic Acid (C)	2.316	200
Nicotinic Acid (B_3)	2.719	260
Pyridoxine Hydrochloride (B_6)	2.873	290
Pyridoxal Hydrochloride (B_6)	3.098	290
Nicotinamide (B_6)	4.838	260
Pantothenic Acid (B_5)	6.293	200
Folic Acid (B_9)	9.247	200
Cyanocobalamin (B_{12})	9.396	200
Riboflavin (B_2)	10.154	290
Biotin (B_7)	10.256	200

4.3.9 Quantification of Metabolites in Spent Media

The extracellular concentrations of lactate, glucose and ammonia were measured using a benchtop BioProfile Flex 2 from Nova Biomedical. This technique was much simpler than HPLC and required no additional preparation of solutions. The machine was calibrated beforehand using the built-in software command, and Quality Control (QC) levels 1-5 from Nova Biomedical were run to ensure measurement accuracy.

4.3.10 Statistical Analysis

The results in this chapter depict the mean value for each data point across the four replicate flasks, and the error bars depict the standard deviation for each mean. The means and standard deviations were calculated using the AVERAGE() and STDEV() functions in Excel, respectively. Brackets with a single asterisk signify clusters of data points that are not significantly different from one another, as determined by single-factor ANOVA tests conducted in Excel ($p < 0.05$).

4.4 Results

4.4.1 Viable Cell Density and ATP per Viable Cell

Figure 4.1 visualizes the viable cell density and ATP per viable cell for each of the four batch cultures. The maximum observed cell density was slightly greater than 7 x 10^6 cells/mL and the culture longevity was 10 days (240 hours). There is very low growth during the first 24 hours, indicating the lag phase. This was followed by an exponential growth phase between 24 and 96 hours and a sudden onset of stationary growth at 96 hours, which lasted until 192 hours. Viability swiftly plummeted during the final two days of the culture.

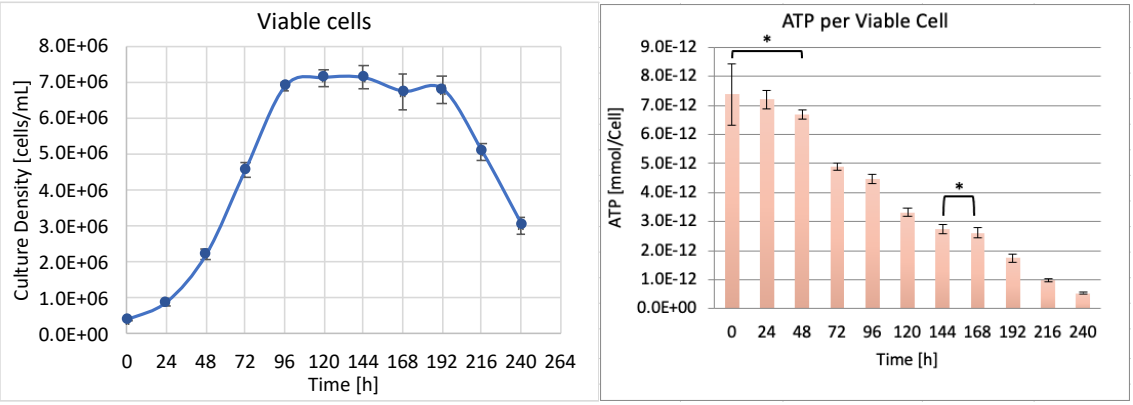


Figure 4.1: Viable Cell Density and ATP Per Cell for GS-CHO Batch Cultures

ATP levels were not significantly different during the first 48 hours, but there was a clear and sudden decline at 72 hours. A steady decline in ATP availability per viable cell was observed for nearly every sequential day following 72 hours, indicating consistently declining ATP availability.

4.4.2 HPLC Quantification of mAb Concentration

Figure 4.2. visualizes the HPLC mAb chromatographs of days 0 and 10 for one flask, showing clearly the production of the antibody by day 10. Differences in concentration between each consecutive pair of samples was calculated and normalized per viable cell to show that specific productivity decreased over time in Figure 4.3. The production of mAb per viable cell did not significantly change during cellular growth but suddenly and progressively declined during the stationary phase and subsequent death.

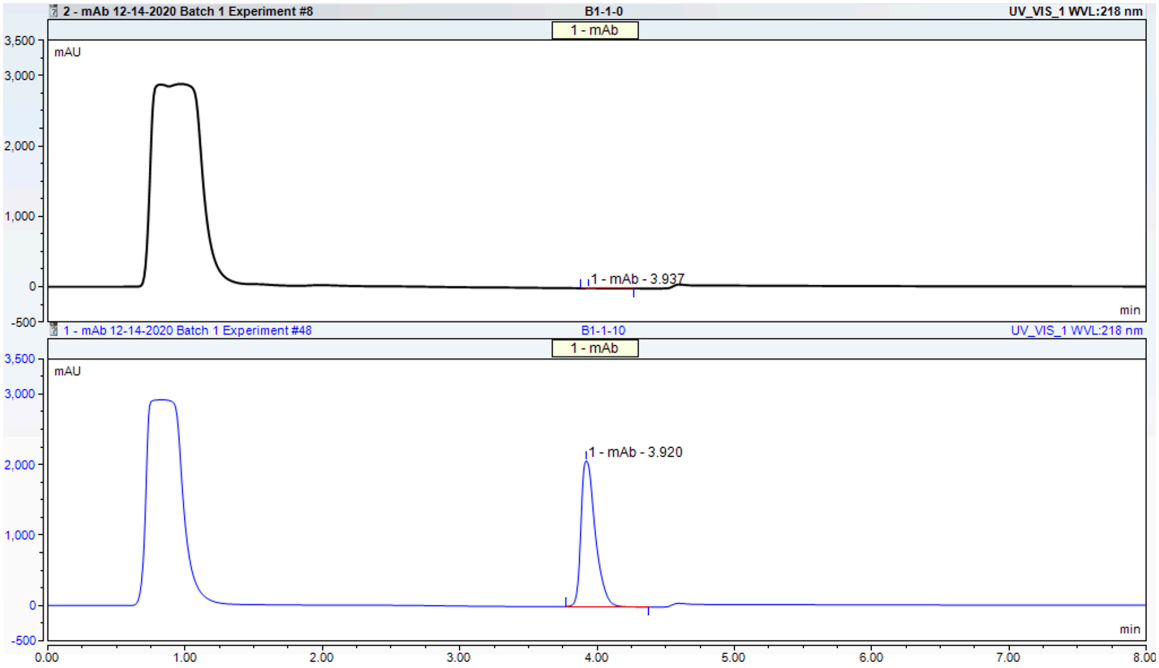


Figure 4.2: HPLC Quantification of mAb Production in Supernatant Samples

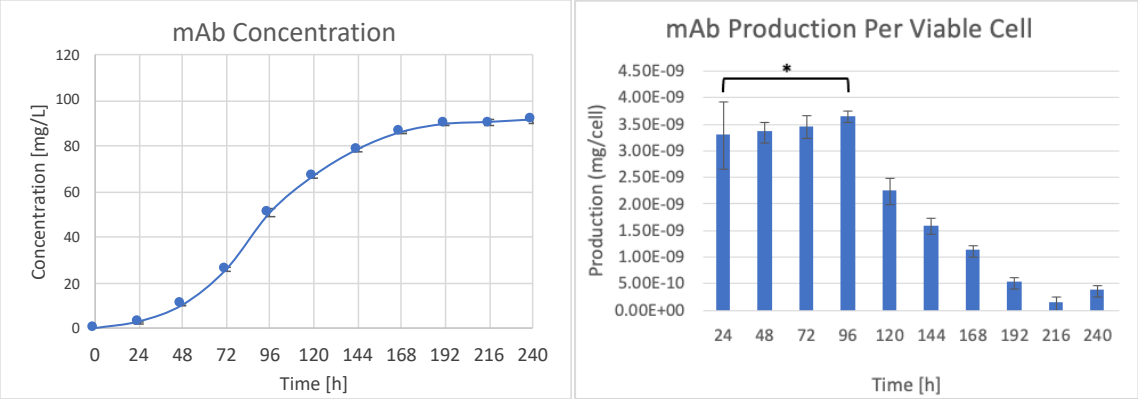


Figure 4.3: mAb Concentration and Production in Supernatant Samples

4.4.3 HPLC Quantification of Amino Acids

Examples of two amino acids HPLC chromatographs of samples taken at day 0 and day 10 for one flask are shown in Figure 4.4 below, visualizing the depletion of select amino acids (such as glutamate, aspartate and asparagine) over the course of the culture. The concentration profiles for key amino acids will be shown and discussed in Chapter 5 within the context of the multivariate analysis. Glutamine, glycine, and alanine (peaks 5, 7 and 10) are not present in the supernatant at the start of the culture because they are initially produced rather than consumed by the cells and are therefore not included in the basal media composition. The amino acids which correspond to the peak windows listed along the top of each chromatograph are given in Table 4.1, and peaks are also labeled by amino acid and retention time on the chromatographs themselves.

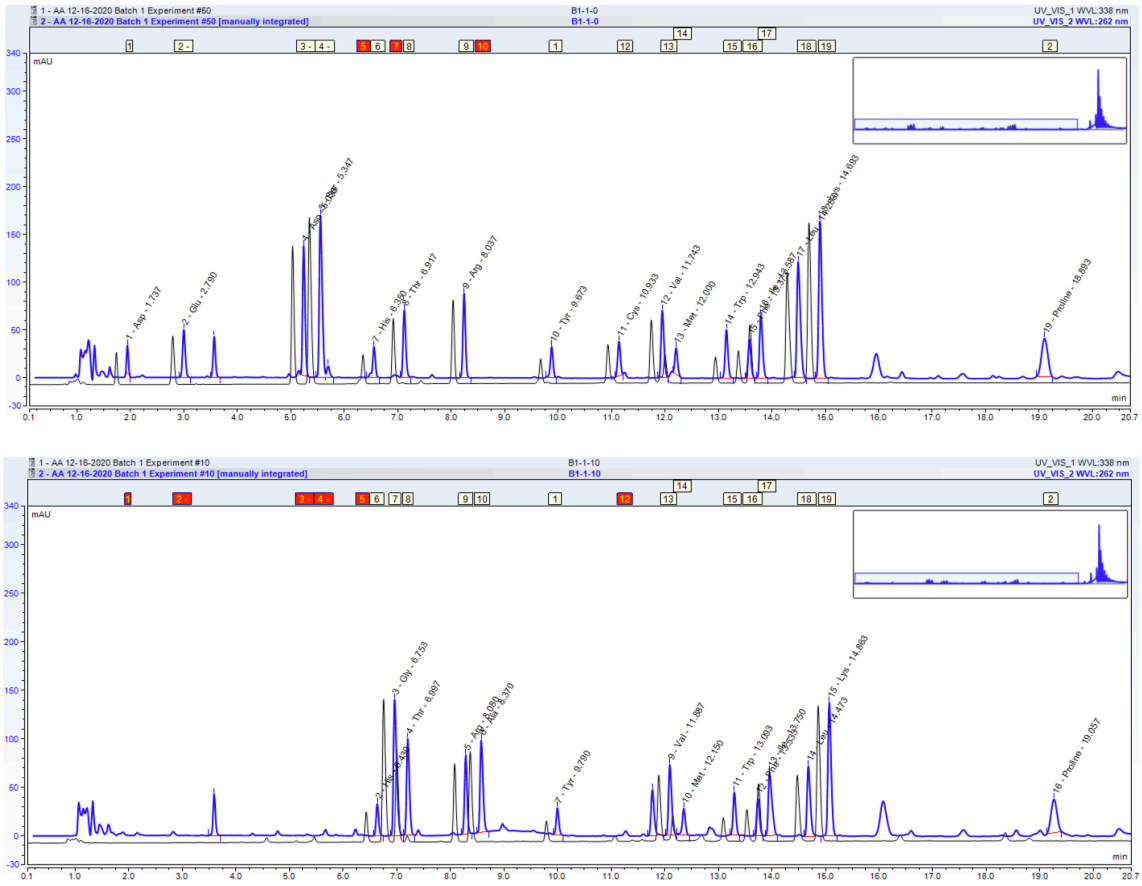


Figure 4.4: HPLC Quantification of Amino Acids in Spent Media (Day 0 and Day 10)

4.4.4 HPLC Quantification of Water-Soluble Vitamins

The chromatographs for days 0 and 10 for one flask are shown in Figure 4.5. All five collected wavelengths are overlaid, but the automatic peak identification based on the defined windows can only be applied to a single wavelength, therefore the vitamins quantified at 200 nm are showed only (the retention times and respective detection wavelengths for the remaining vitamins are identified in table 4.2).

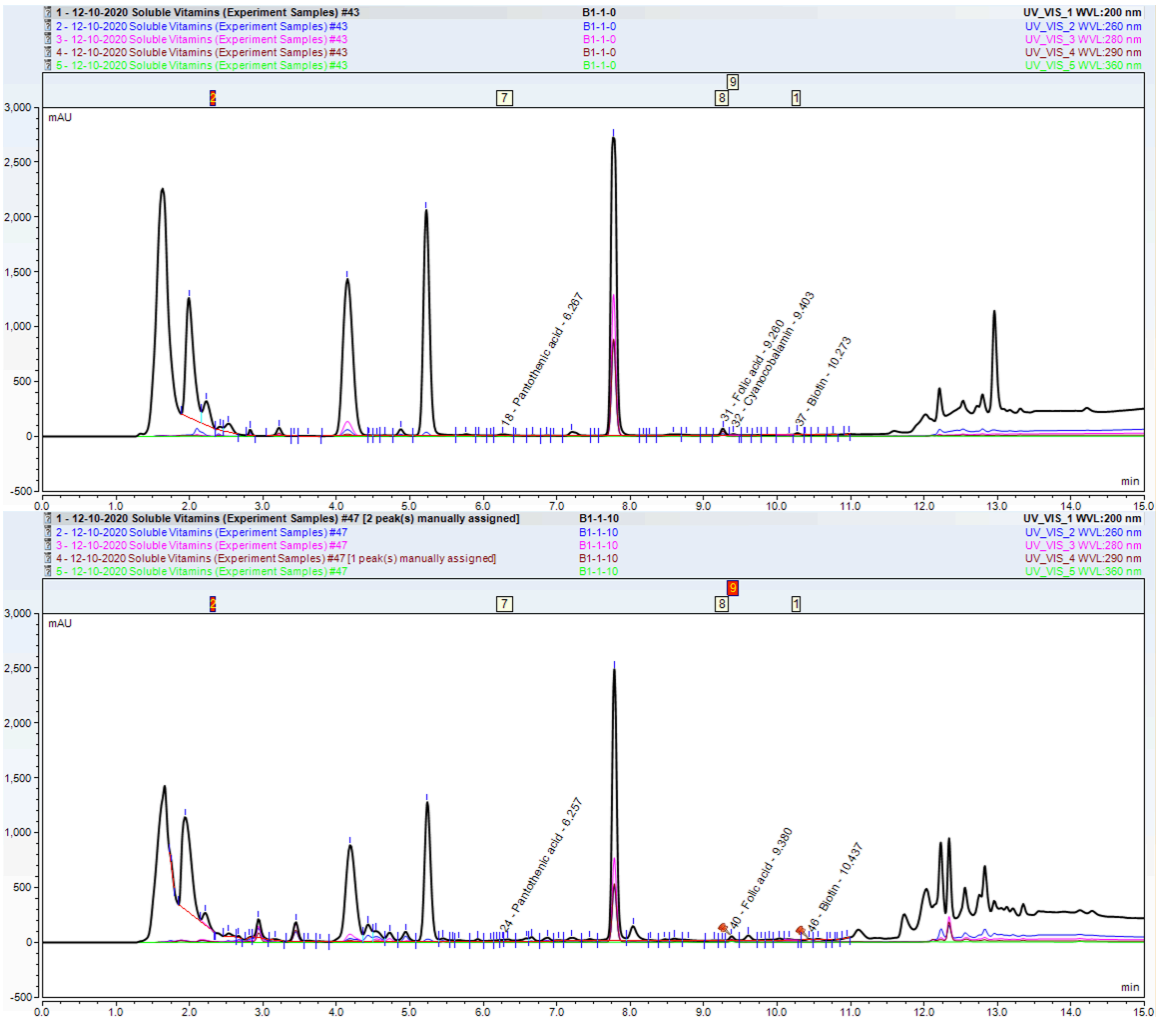


Figure 4.5: HPLC Quantification of Vitamins in Spent Media (Day 0 and Day 10)

4.4.5 Glucose, Lactate, and Ammonia Consumption and Production

Graphs of specific lactate and ammonia *production* are depicted in Figure 4.6 below their respective graphs of extracellular concentrations, since these metabolites were primarily produced. Glucose, however, was exclusively consumed, so a graph of specific *consumption* was chosen instead. Glucose was steadily depleted until 192 hours, when the extracellular concentration reached approximately 0 mM. Ammonia, a common byproduct of amino acid metabolism, steadily accumulated except for a temporary stall in production

between 96 and 144 hours. This shift correlated strongly with the decrease in specific production of lactate that preceded its eventual consumption. In general, it can be seen that the specific consumption and production of these metabolites was highest during the lag and exponential phases and stabilized at lower magnitudes during the stationary phase.

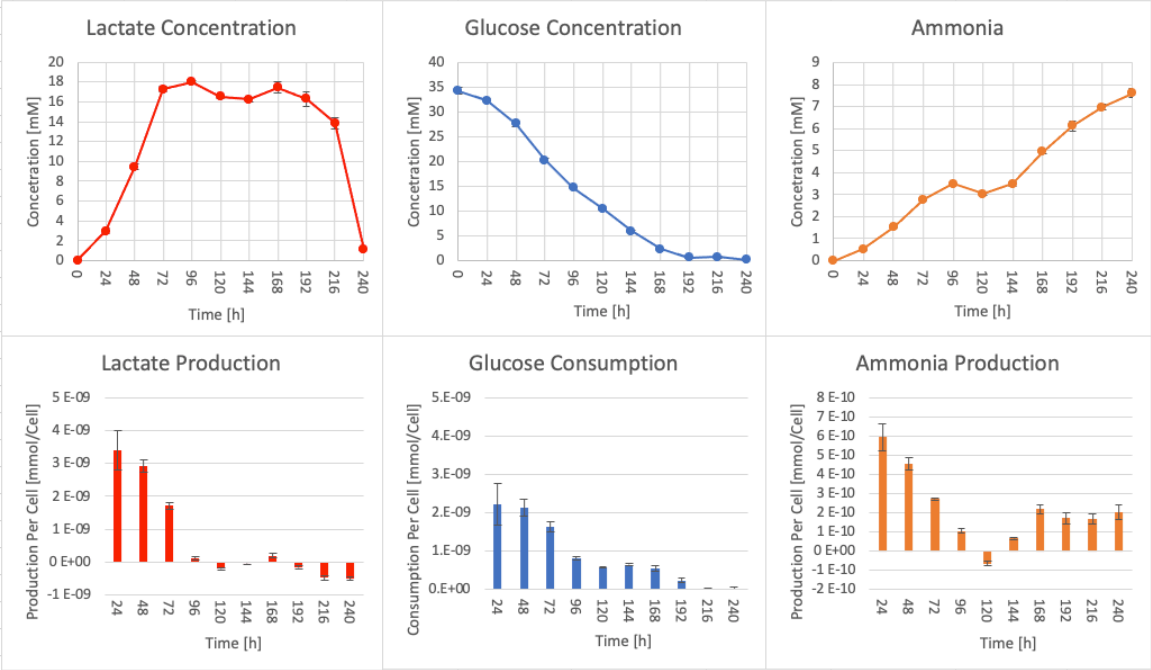


Figure 4.6: Extracellular Concentrations and Consumption/Production of Metabolites

4.5 Discussion

4.5.1 Growth, Viability and mAb Productivity

The observed growth and viability trends for these cultures were as expected – a brief lag in growth during the first 24 hours, followed by very clearly defined growth, stationary and death phases. The consistently declining availability of ATP per cell was also predictable, given that the cells were not fed any supplemental nutrients during the culture to produce increased amounts of ATP. The mAb productivity was rather low at a maximum titer of

under 100 mg/L, and one possible reason for this is the multiple occurrences of liquid nitrogen exhaustion in our cryopreservation tank (presumably due to a leak) which resulted in compromised viability of the existing vials. The productivity of the recovered cells may have been affected by these incidents, but it is also possible that these cells did not originate from a very productive clone to begin with. Regardless, the low productivity of the cells may not be generally ideal but is acceptable for the purposes of these experiments because it theoretically leaves more room for future optimization.

4.5.2 HPLC Quantification of Amino Acids and Water-Soluble Vitamins

The method for HPLC quantification of amino acids was robust and successful, and the key results from the analysis will be shown and discussed in greater detail in the following chapter. Unfortunately, the vitamins method was overall less robust than the amino acids quantification method, as the baseline was unstable in certain segments and the resolution of some peaks was very poor (as can be seen in Figure 4.5). This resulted in the need for manual integration when peaks were unable to be correctly identified by Chromeleon. This inevitably establishes some potential for error in the calculations for concentration, so there is somewhat less confidence in the accuracy of these results compared to the amino acid concentrations. Future changes will need to be explored in the optimization of this protocol, such as alternative mobile phases, phase gradients, flow rates and sample preparation.

4.5.3 Glucose, Lactate and Ammonia

The trends observed for glucose, lactate and ammonia in the BioProfile analysis depicted in Figure 4.6 are typical for mammalian cell culture – a halt in the accumulation of lactate

can be seen just before the onset of stationary growth, and this shift is documented very extensively in the literature (although still not fully understood) [62]. A corresponding decrease in the specific consumption of glucose is also observed alongside this shift due to an increased dependence on anaplerotic pathways in the TCA cycle. These findings will be juxtaposed alongside the profiles for key amino acids in Chapter 5 where the clusters and shifts identified in the multivariate analysis will be explored in close detail to hypothesize causational relationships between the depletion and production of metabolites and growth behavior.

CHAPTER 5

MULTIVARIATE ANALYSIS OF GS-CHO BATCH CULTURE DATA

5.1 Introduction

An understanding of energy metabolism in mammalian cells is critical for the optimization of media composition. Culture media provides extracellular nutrients such as amino acids, glucose, lipids, vitamins, inorganic salts and trace metals which are necessary to support metabolic processes for cell survival, growth and proliferation, biomass composition, and recombinant protein/antibody production [34]. Vitamins cannot be endogenously produced but play important roles as enzyme cofactors and as antioxidants in redox metabolism (as do some trace metals) [41], [42]. Lipids are important for plasma membrane composition and lipid metabolism, and amino acids and glucose significantly contribute to energy production. Glucose is a 6-carbon sugar that is broken down in the cytosol into two 3-carbon molecules of pyruvate in an aerobic process known as glycolysis, which simultaneously produces two ATP and NADH molecules each from the reduction of NAD^+, shown at the top of Figure 5.1.

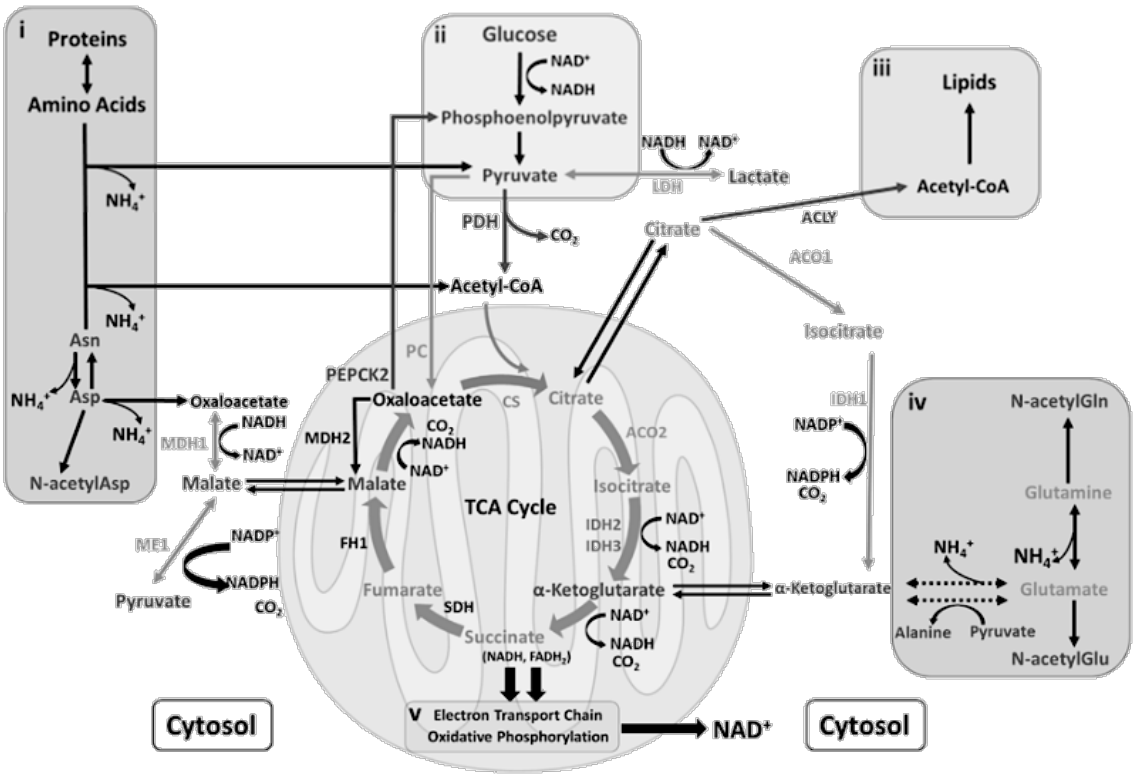


Figure 5.1 High-Level Overview of Energy Metabolism (used with permission from [65])

Pyruvate can then be anaerobically metabolized to oxidize NADH into NAD^+ (for the continuation of glycolysis), which produces the toxic metabolite lactate in a process known as lactic acid fermentation. Alternatively, in the presence of enough oxygen, pyruvate is instead converted into acetyl-CoA by pyruvate dehydrogenase (PDH) as the first step in the tricarboxylic acid (TCA) cycle. The TCA cycle is an aerobic alternative to lactic acid fermentation that takes place in the mitochondria and generates one GTP, three NADH, one $FADH_2$ and two CO_2 molecules in a single cycle. It involves many intermediates, as shown in Figure 5.1, which can be synthesized from free amino acids by transaminase enzymes; these alternative synthesis routes are formally referred to as anaplerotic pathways, and their existence dictates a significant necessity for amino acid availability in media composition.

Additionally, it should be noted that these pathways contribute substantially to ammonia production and accumulation, as can be seen in the reactions for glutamine and asparagine conversion into glutamate and aspartate, respectfully (as well as many others not explicitly shown in the figure). A summary of the amino acids involved in anaplerotic pathways for the TCA cycle is given in Table 5.1, adapted from findings reported in a biochemistry textbook published in 2020 [66].

Table 5.1: Anaplerotic Pathways for TCA Cycle Precursors and Intermediates [66]

Amino Acid Precursors	TCA Cycle Precursors/Intermediates
Glycine → Serine Methionine → Cysteine Tryptophan → Alanine	Pyruvate
Phenylalanine → Tyrosine Tryptophan Leucine Isoleucine Lysine	Acetyl-CoA
Arginine → Glutamate Proline → Glutamate Histidine → Glutamate Glutamine → Glutamate	A-Ketoglutarate
Valine Isoleucine Methionine Threonine	Succinyl-CoA
Asparagine → Aspartate Phenylalanine → Tyrosine	Fumarate
Asparagine → Aspartate	Oxaloacetate

The electron-carrying coenzymes produced by the TCA cycle, NADH and $FADH_2$, then proceed to the electron transport chain where they undergo a series of redox reactions to

generate a significantly greater amount of ATP than with glycolysis alone. Upregulated TCA cycle activity is, however, associated with redox stress and increased generation of reactive oxygen species (ROS), which must be neutralized by antioxidant molecules. The shifting dependencies on and away from lactic acid fermentation and the TCA cycle throughout mammalian cell culture, and the metabolic behaviors that characterize such activity, are a significant research area for ongoing studies. These principles of energy metabolism are important to understand as a basis for drawing conclusions from the data collected in the GS-CHO cultures.

5.2 Aims and Objectives

The aim of this chapter is to apply the multivariate analysis platform previously established to conglomerate the data and profile cellular behavior at a high level. Once significant shifts and clusters of similarity are identified, knowledge of energy metabolism will be used to form hypotheses on the shifting metabolic behavior of the cultures and identify areas in need of optimization. From the metabolic profiles data, specific consumption will be calculated to analyze metabolic behavior from a different perspective. Given the compounded error from the specific consumption calculations (standard deviations in both viable cell density and the metabolite consumption), single-factor ANOVA will be performed to elucidate only truly significant shifts in consumption patterns for each variable and additional multivariate analysis will be performed which reflects this.

5.3 Materials and Methods

5.3.1 Creation of an Excel Template File for Data Consolidation

To streamline the data input process and ensure accurate reproducibility, an Excel template file was created with the proper formatting to allow for simple copying and pasting from instrument reports. The export formatting of each instrument report, as well as the retention times of everything being measured (in the case of HPLC) was used to guide the formatting of the Excel template to maximize ease of use. The template was designed such that raw data from each report would be converted to units of mM (if necessary), averaged across replicates for the calculation of mean and standard deviation, and graphs for each data component would be automatically populated with pre-formatted axis labels and standard deviation bars. Different graphs and calculations for different analyses were segregated by sheet, and all final data was consolidated into a separate sheet which could be exported in .CSV file format for multivariate analysis in R. The dataset analyzed was composed of 11 averaged observations of 40 independent variables: all 20 amino acids (mM), 11 water soluble vitamins (mM), glucose (mM), lactate (mM), ammonia (mM), ATP per cell (mmol/cell), mAb titer (mg/L), and viable/apoptotic/dead/total cell densities (cells/L).

5.3.2 Multivariate Analysis

The HCA and PCA were conducted using the same R code and data processing as in Chapter 3, and the same correlation between culture growth stages and identified clusters is applied. Both vertical and horizontal dendrograms are applied to the HCA heatmaps for clustering analysis, but the vertical clustering is removed in additional figures to ensure temporal ordering of data points, which improves visual clarity of metabolic shifts.

5.3.3 Statistical Analysis

In addition to the mean and standard deviation calculations for the metabolite profiles performed previously, additional single-factor ANOVA tests were applied to the calculated specific consumption data. These ANOVA tests were performed using Excel macros which automatically executed the single-factor ANOVA function in the Analysis Toolpak against sequential pairs of data (for example: 0-24 and 24-48 hours, 24-48 and 48-72 hours, etc.). The specific consumption between two pairs of data was deemed to be significantly different only if $p < 0.01$ in order to highlight only the most significant metabolic shifts for clarity of visualization. Insignificantly different clusters of data were averaged using the mean of all values in the cluster and the values for each data point within clusters was replaced by their respective average. A new .csv file was created in Excel to reflect these changes and subsequently analyzed in R for HCA and PCA.

5.4 Results

5.4.1 Hierarchical Clustering Analysis of Concentration Profiles

The HCA performed on extracellular concentration profiles is shown in Figure 5.2. Data were standardized using the z-scoring methodology previously described. It can once again be seen that each of the four stages of growth exhibits a unique metabolic profile, as shown by the unsupervised formation of four clusters in Figure 5.2 which have a strong correlation with the time points for each growth phase seen in the viable cell density graph (Figure 4.1).

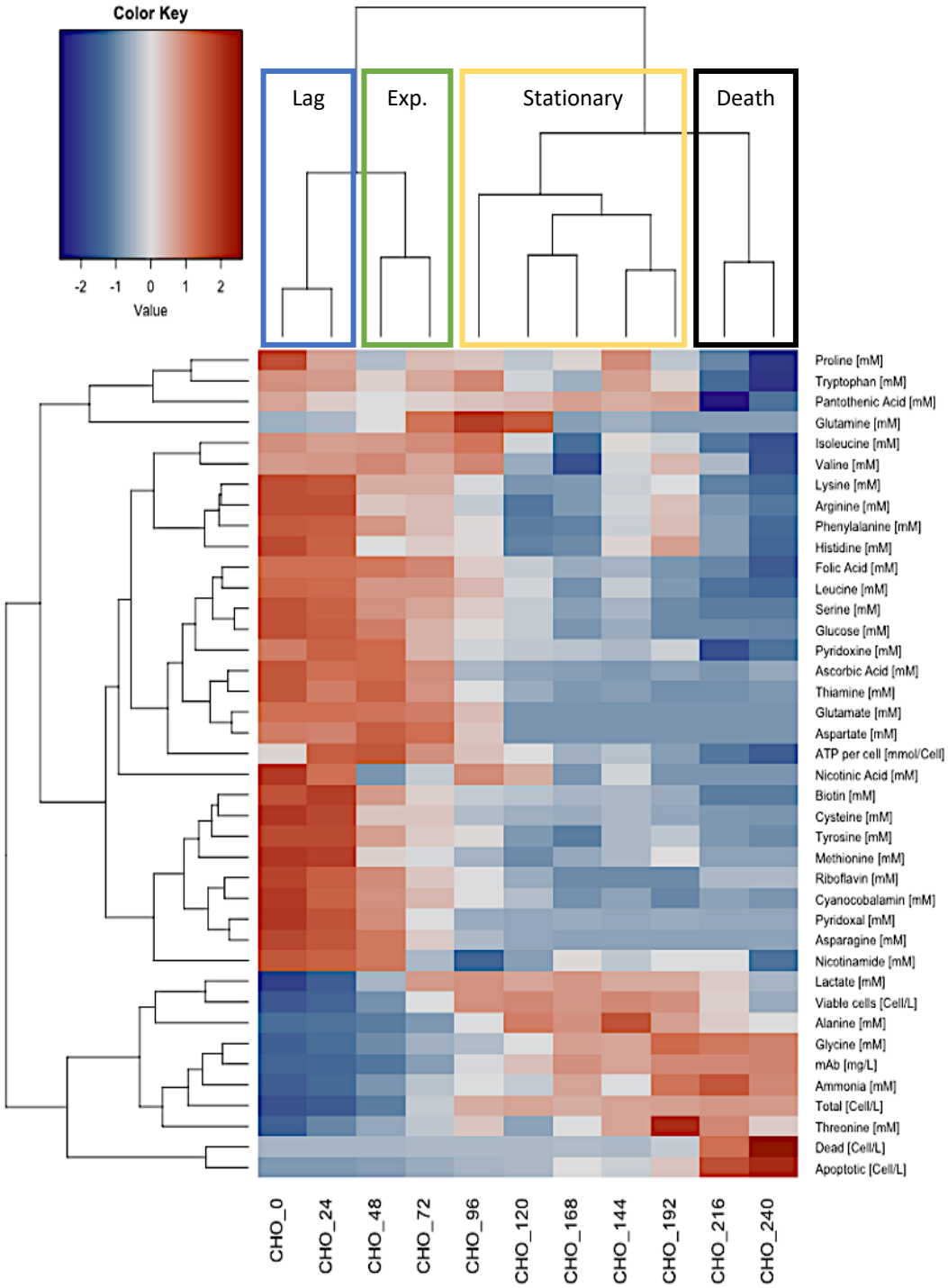


Figure 5.2 HCA of GS-CHO Batch Culture

Additionally, a high-level view of the metabolic shifts can be seen – particularly from exponential to stationary phase. It is at this point that a majority of vitamins and amino acids

exhibit a significantly lower extracellular concentration compared to the population mean (as indicated by the sudden shift from red to blue). The bottom quarter of the heatmap shows the cluster of variables that yielded an increased concentration over time, namely toxic metabolites (lactate, ammonia), mAb titer, dead/apoptotic cells, and a few amino acids (threonine, glycine and alanine). Vertically, the profile of viable cell density is clustered the closest with lactate because the initial shift away from lactate production correlates exactly with the onset of the stationary growth phase, and the rapid decline in viable cell density coincides with significant lactate consumption. In terms of amino acids that were observed to accumulate in the supernatant, glycine was steadily produced throughout the time course of the culture and there was a notable shift from production to consumption of alanine at 144 hours. While threonine production was observed in the HCA, further inspection of its concentration profile led to the conclusion that the actual increase was not very significant when accounting for standard deviation, and the overall variation in its concentration was small in magnitude, so this was not explored further. The graphs of concentration profiles for these three amino acids are depicted in Figure 5.3.

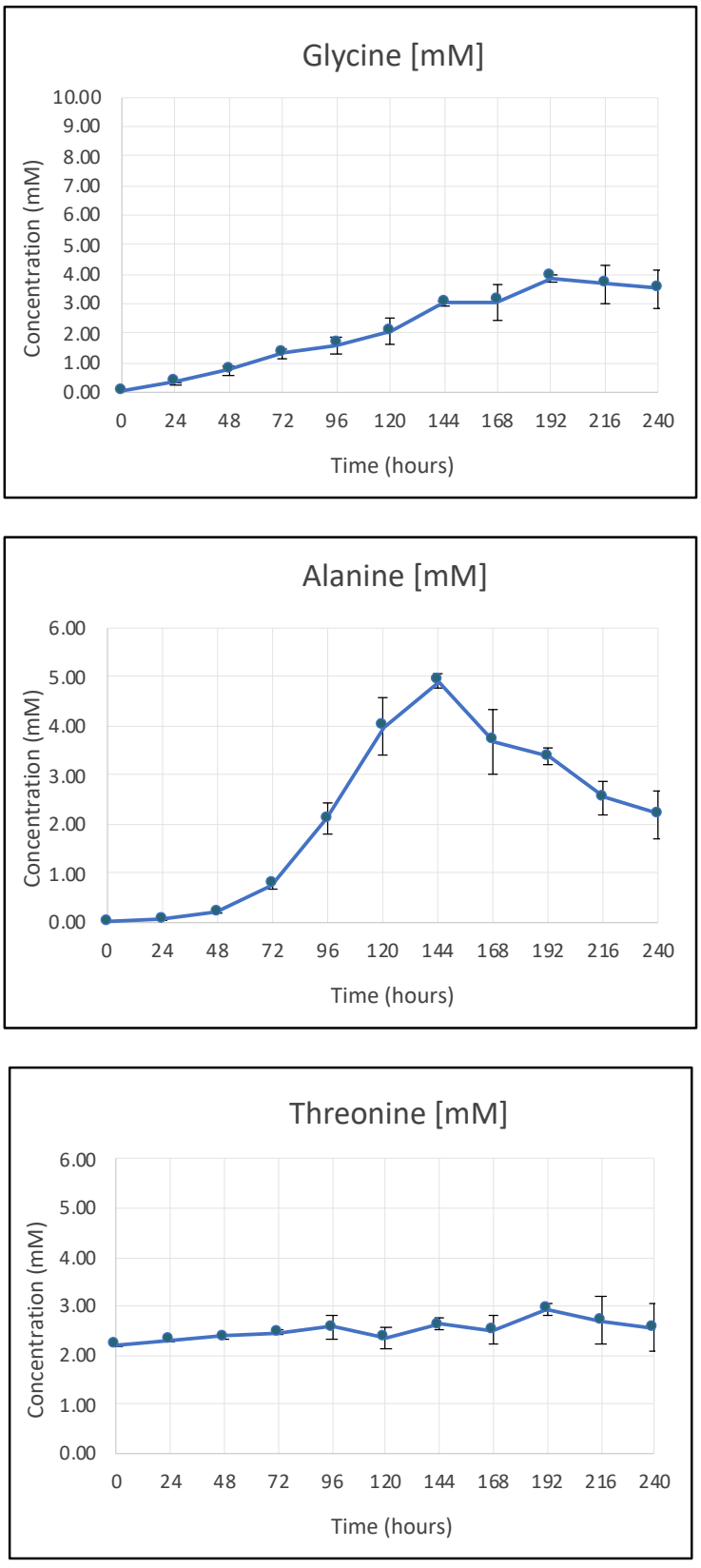


Figure 5.3: Extracellular Profiles of Glycine, Alanine and Threonine in CHO Culture

Glutamine appears to exhibit consistent production until 96-120 hours, after which it experiences consumption. This shift was observed to immediately follow the exhaustion of glutamate at 120 hours. Directly following the exhaustion of glutamine at 144 hours, it can be seen that alanine shifted away from production immediately towards consumption. Following glutamate exhaustion, it was observed that ammonia production began once again following a brief period of consumption between 96 and 120 hours. A graph of this relationship is depicted below in Figure 5.4:

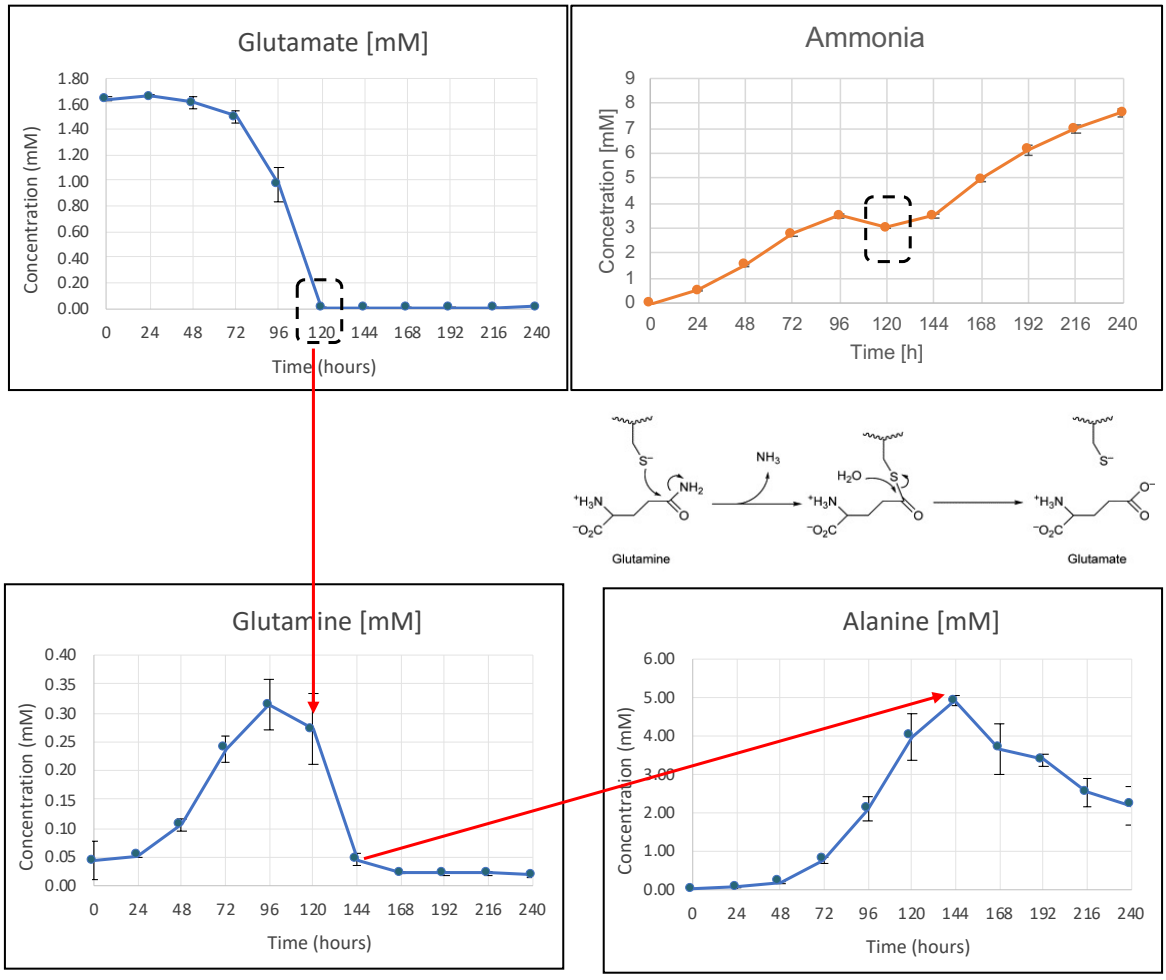


Figure 5.4: Glutamate, Glutamine and Alanine Metabolic Shifts

Aspartate was clustered most closely with glutamate in the HCA, and upon closer inspection, it was discovered that this was because it too was completely exhausted by 120 hours. The concentration profiles of aspartate and asparagine, which are related through their relationship with oxaloacetate in the TCA cycle, are shown in Figure 5.5 below; both are nearly fully exhausted by 120 hours.

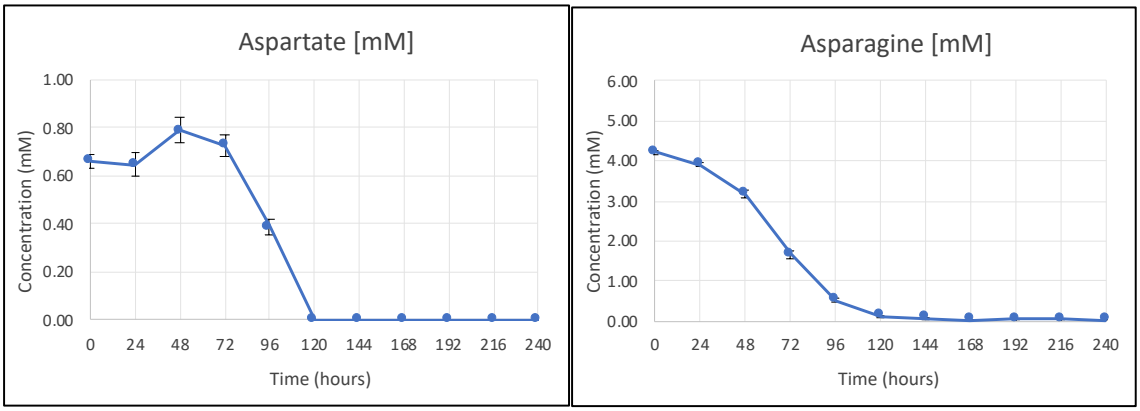
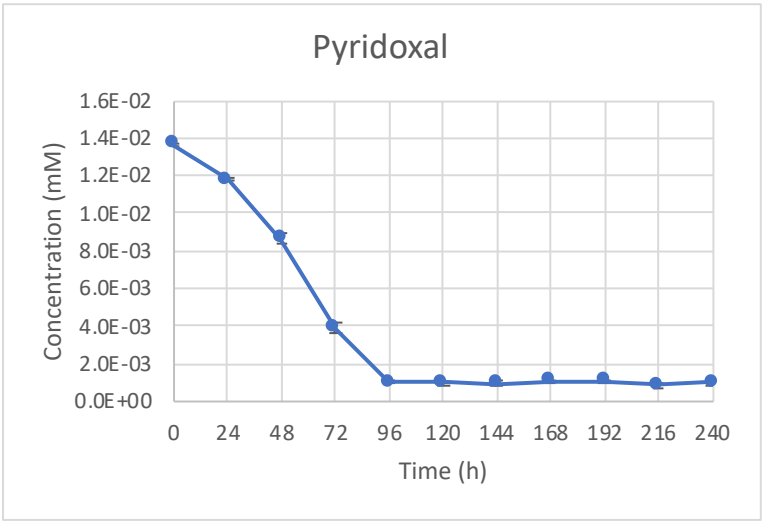


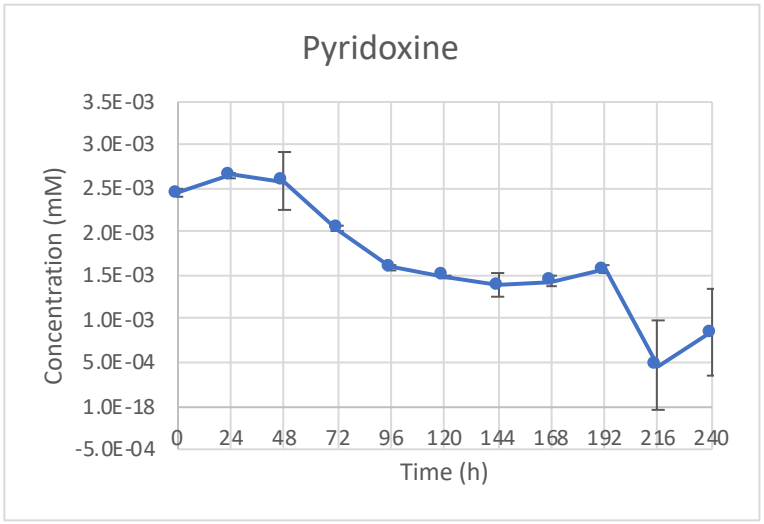
Figure 5.5: Extracellular Aspartate and Asparagine Exhaustion

Valine, arginine, phenylalanine and histidine exhibited slightly higher extracellular concentrations at 192 hours compared to the neighboring time points in the HCA, but upon further inspection of their independently graphed profiles, these changes were not very significant statistically nor were they of notably high magnitude. Therefore, this observation was not investigated further.

All vitamins were consumed in culture to an extent; of those measured, pantothenic acid conveyed the most striking change in consumption pattern. A steady concentration was observed up until 216 hours when a sharp drop was discernable, and despite insignificant consumption prior to this point, pyridoxine exhibited a similar drop in concentration

at 216 hours. Pyridoxal, another vitamin in the B_6 family with functionally similar significance to pyridoxine, was exhausted around 96 hours – differing greatly from the pattern observed for pyridoxine. Most remaining vitamins showed shifts in consumption that coincide with the onset of the stationary growth phase at 96 hours and the relatively decreased metabolic activity associated with it. The concentration profiles for pyridoxal, pyridoxine and pantothenic acid are shown in Figure 5.6.





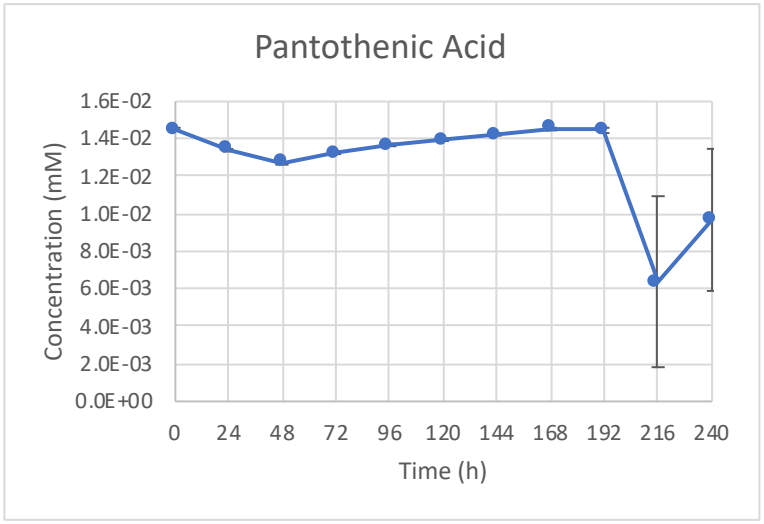


Figure 5.6: Extracellular Concentrations of Pyridoxal, Pyridoxine and Pantothenic Acid

5.4.2 Principal Component Analysis of Concentration Profiles

PCA was performed on the data and visualized in Figure 5.7. The first principal component accounts for a staggering 73.47% of the total variation, and PC2 is far less significant at 11.82%. The distinct metabolic profile of each growth phase can be seen by the clearly separated clusters of data points, color-coded to coordinate with their respective previously identified HCA cluster. The first two data points are spatially proximal, indicating smaller change in extracellular concentrations likely as a result of initial low cell density. As the cells enter exponential growth, it can be seen that metabolic activity is high as the concentration profiles change significantly during each sample interval. After the onset of the stationary phase at 96 hours, the metabolic changes are less significant because viable cell density is maintained at steady state. The last two data points are significantly separated from the other clusters, representing the rapid cell death observed during this point in the culture.

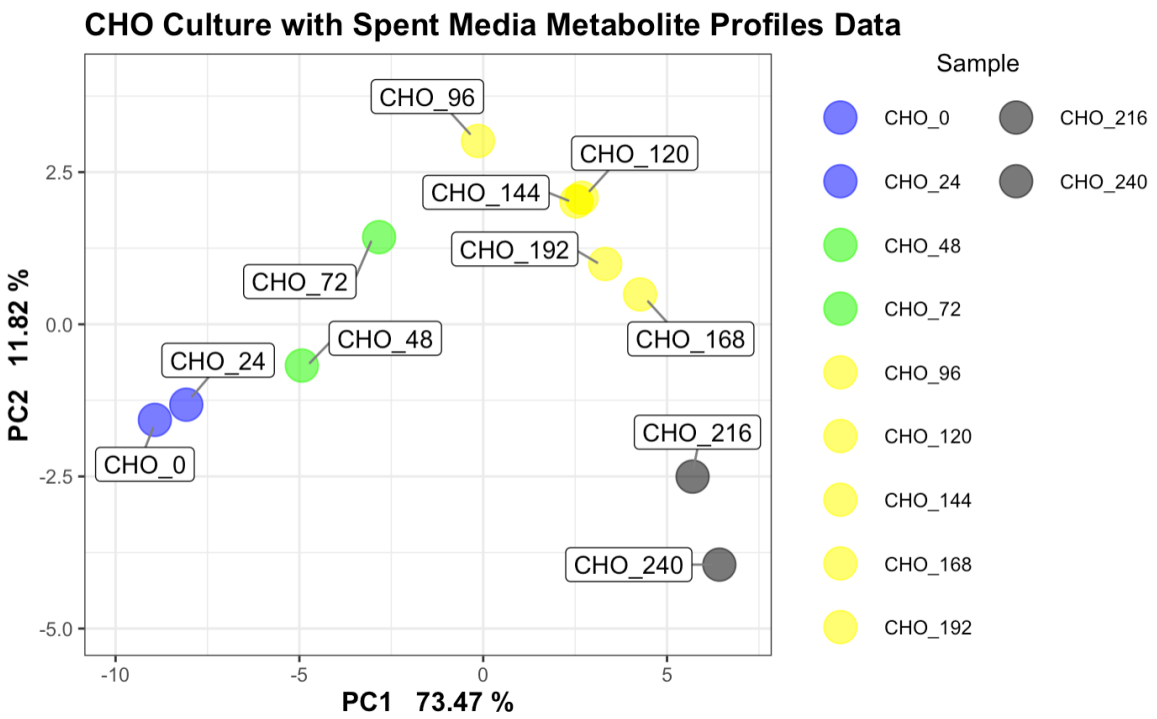


Figure 5.7: PCA of GS-CHO Batch Culture

5.4.3 Multivariate Analysis of Specific Consumption and Production of Nutrients

After comprehensive analysis of the spent media profiles, specific production and consumption were calculated for each variable and analyzed multivariately. Calculations were performed by dividing the difference in extracellular concentration (mM, or mmol/L) between two consecutive data points by the viable cell density (cells/L) recorded at the time of the later datapoint, resulting in a final value of mmol/cell. Positive values indicate net consumption per cell during that interval of time, and negative values indicate net production per cell. This data, however, was not utilized in the multivariate analysis because it was too difficult to discern significant relationships. In order to visualize more meaningful results, a single-factor ANOVA test was run for each consecutive pair of values to identify significant shifts within the data ($p < 0.01$). Clusters of values for each measured variable

that were statistically insignificant from one another between sequential time points were averaged such that all data points within the cluster would exhibit the same value in the HCA heatmap. The resulting HCA and PCA are shown in Figures 5.8 and 5.9, and have substantially enhanced clarity due to the elucidation of significant variation only.

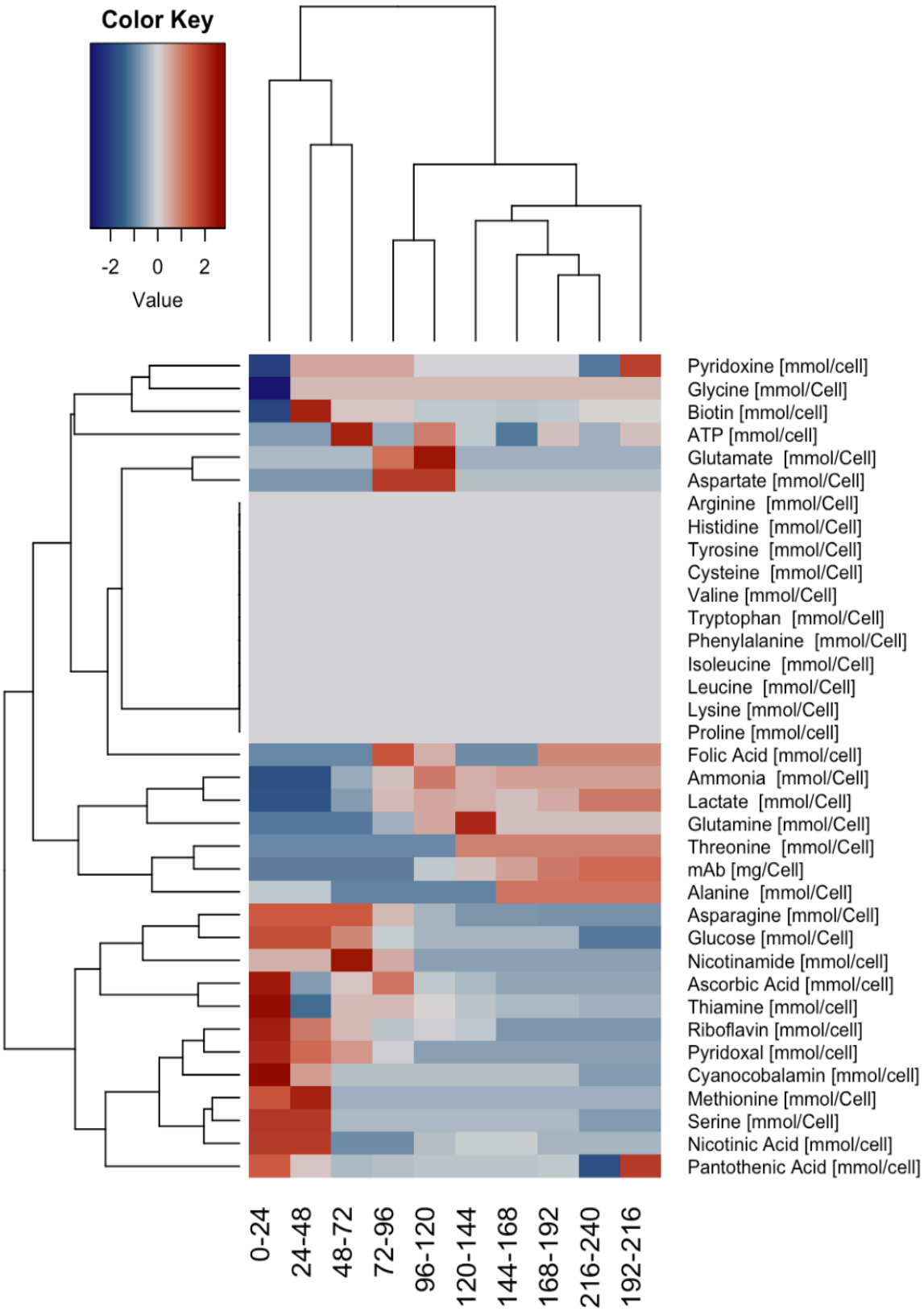


Figure 5.8: HCA of GS-CHO Specific Consumption Data in Batch Culture

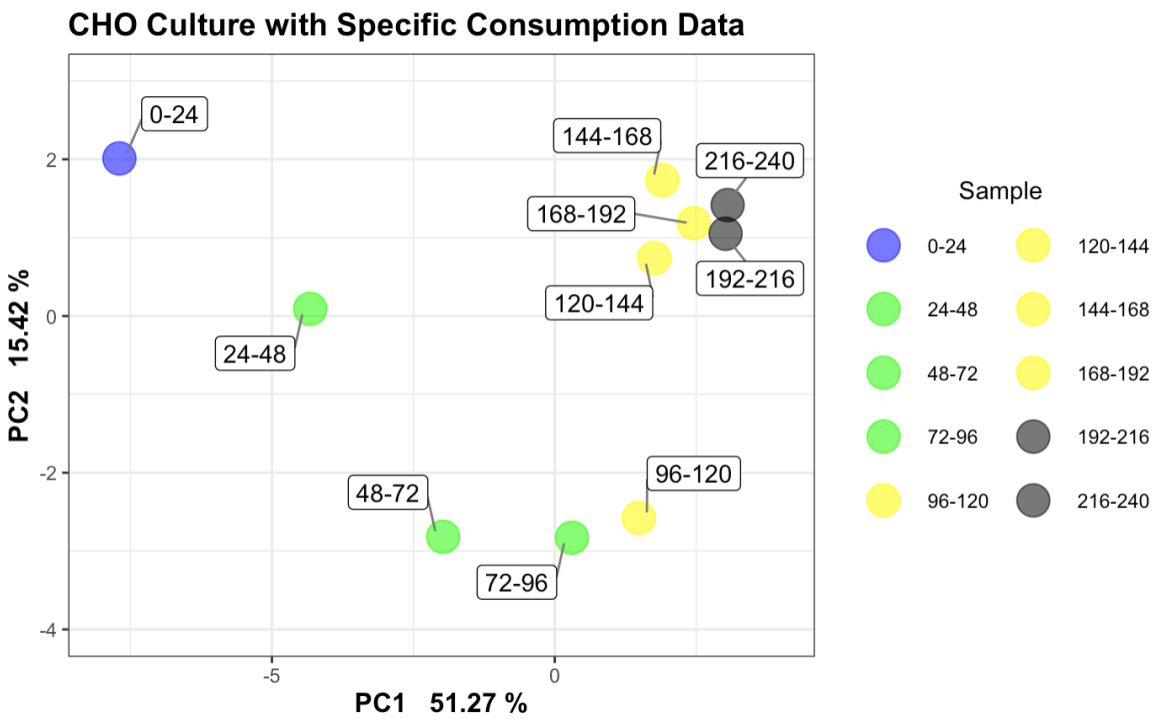


Figure 5.9: PCA of GS-CHO Specific Consumption Data in Batch Culture

Immediately, the differences between the specific consumption data and the previously presented extracellular concentration data are visually obvious. There is a large block of the HCA heatmap which is purely gray, indicating no significant variation ($p < 0.01$) within the entire rows (according to single-factor ANOVA considering all groups). All members of this block are amino acids that were not identified as significant in the prior analysis, so it was concluded that these were not limiting nutrients for the cultures. Another key difference between the HCA for specific consumption and extracellular profiles is the vertical clustering, because the four most significant clusters no longer correlate perfectly with the four stages of culture growth; instead, there is a segmentation of the exponential phase at 72 hours and a distinct cluster between 72 and 120 hours which encompasses the transition from late exponential into stationary. All time intervals after 120 hours are

clustered together, with the interval from 216-240 hours (first interval of rapid cell death) being a standout datapoint within the cluster. From the PCA in Figure 5.9, a clear divide can be drawn distinguishing data points before 120 hours and those taken after 120 hours. The first five days of the culture are characterized by very dynamic specific consumption behavior, suggesting a constantly evolving metabolic state prior to 120 hours. After 120 hours, the specific consumption of media components appears to insignificantly change, as indicated by the comparatively much tighter cluster. In order to better visualize dynamic specific consumption in chronological order, another heatmap was generated without the inclusion of column clustering to allow for the temporally sequential ordering of data points, shown in Figure 5.10.

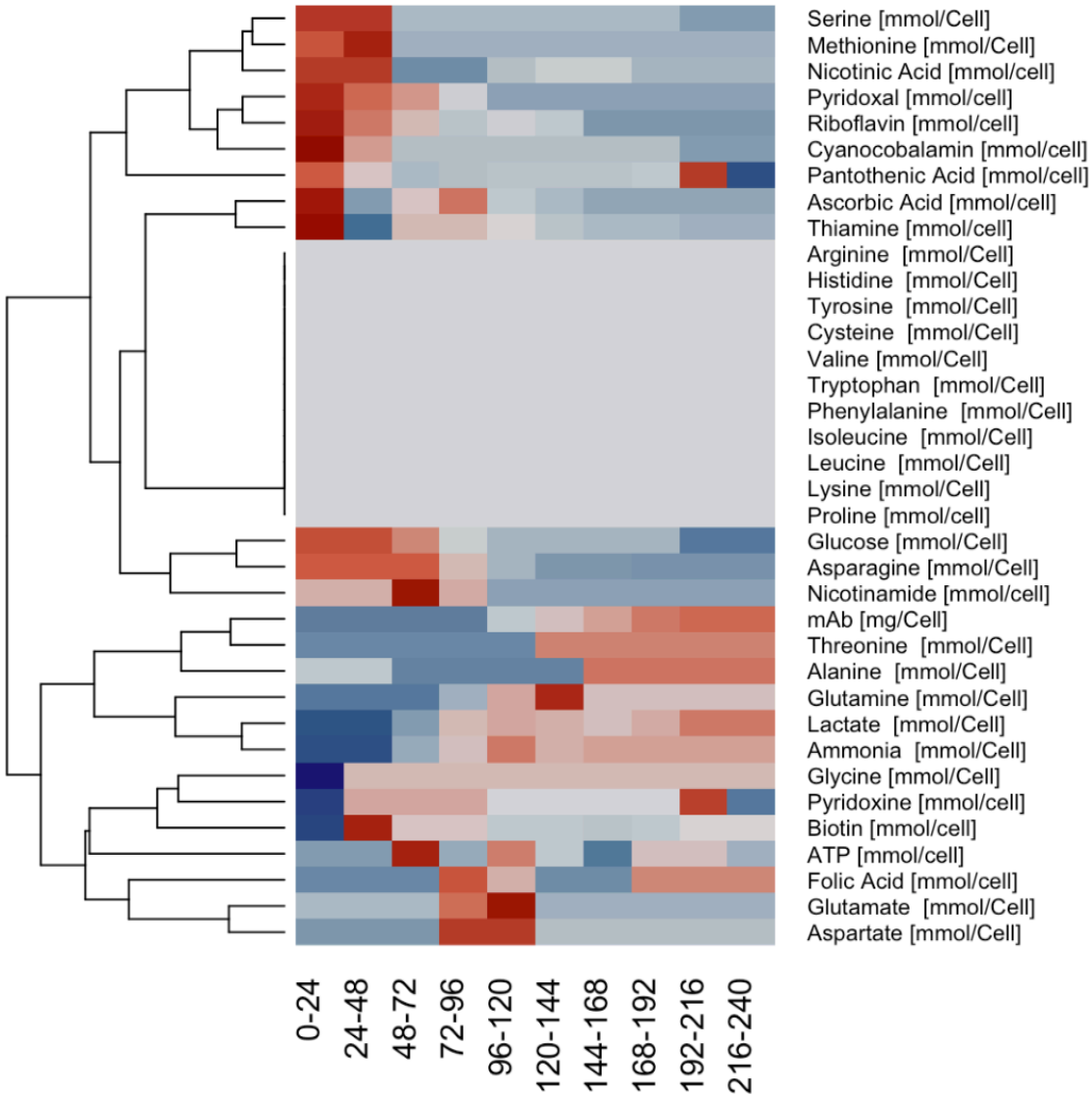
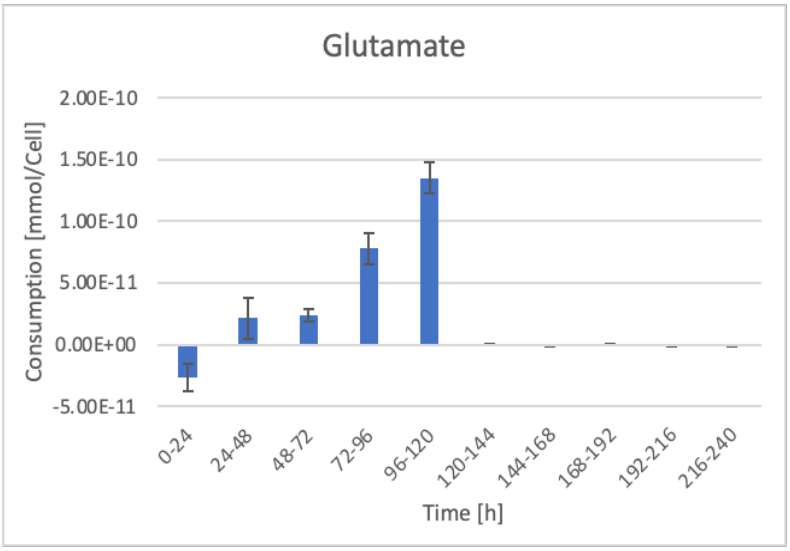
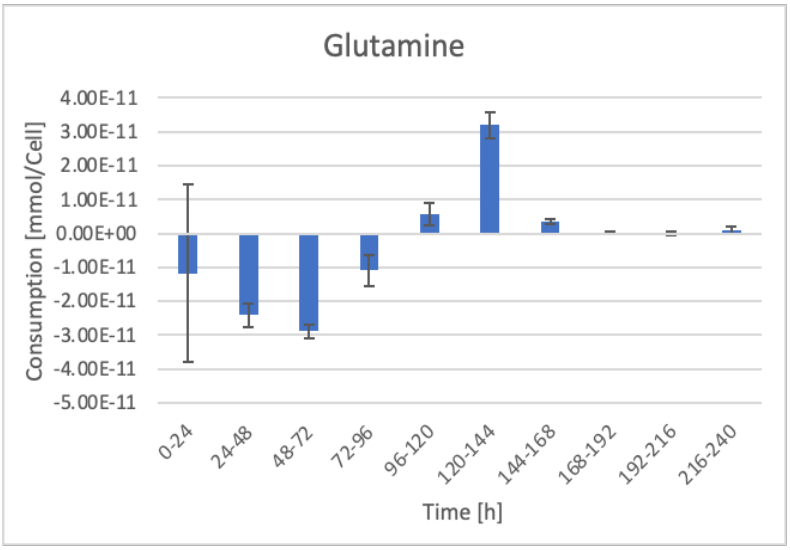


Figure 5.10: HCA of GS-CHO Specific Consumption Data (no column clustering)

A high-level observation that can be made from the heatmap is that a majority of the nutrients exhibit heightened specific consumption during the first 48-72 hours (before late-stage exponential growth). Given the established significance of 120 hours as a point of metabolic shift, as seen in the HCA and particularly PCA, focus was placed on changes that occurred shortly before and after then. It is known from the extracellular profile data that glutamate was completely exhausted at 120 hours, and this is also confirmed by the specific

consumption data as indicated by the dark red square (high consumption) between 96 and 120 hours. It can also clearly be seen that high specific consumption of glutamine between 120 and 144 hours follows glutamate exhaustion, and an immediate shift from consistent production to consistent consumption of alanine follows glutamine exhaustion at 144 hours. The graphs of specific consumption for these three amino acids are depicted in Figure 5.11.





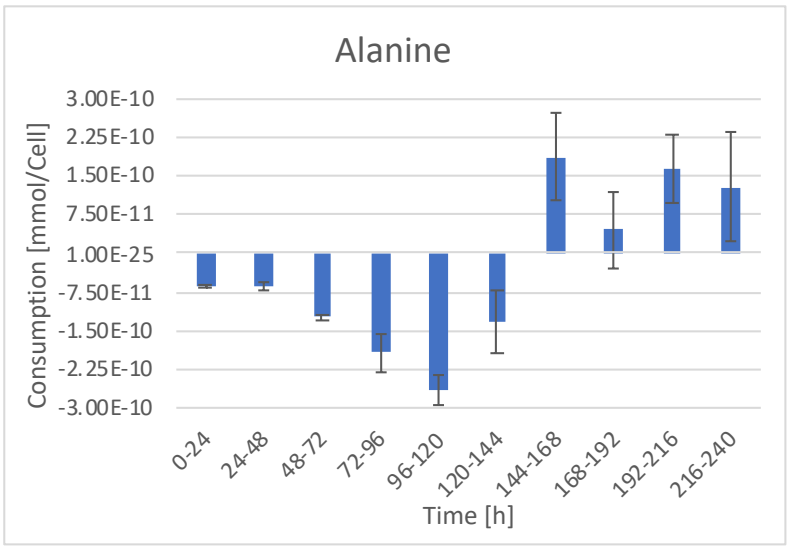


Figure 5.11: Specific Consumption of Glutamate, Glutamine and Alanine

The relationship between glutamate and aspartate consumption is also elucidated more clearly than before, as there is a sudden onset of specific consumption during late-stage exponential growth (72-96 hours) carrying over into early stationary growth, until their mutual exhaustion at 120 hours. Asparagine, the anaplerotic precursor to aspartate, is consumed heavily at the start of the culture until it is exhausted around the same time as aspartate.

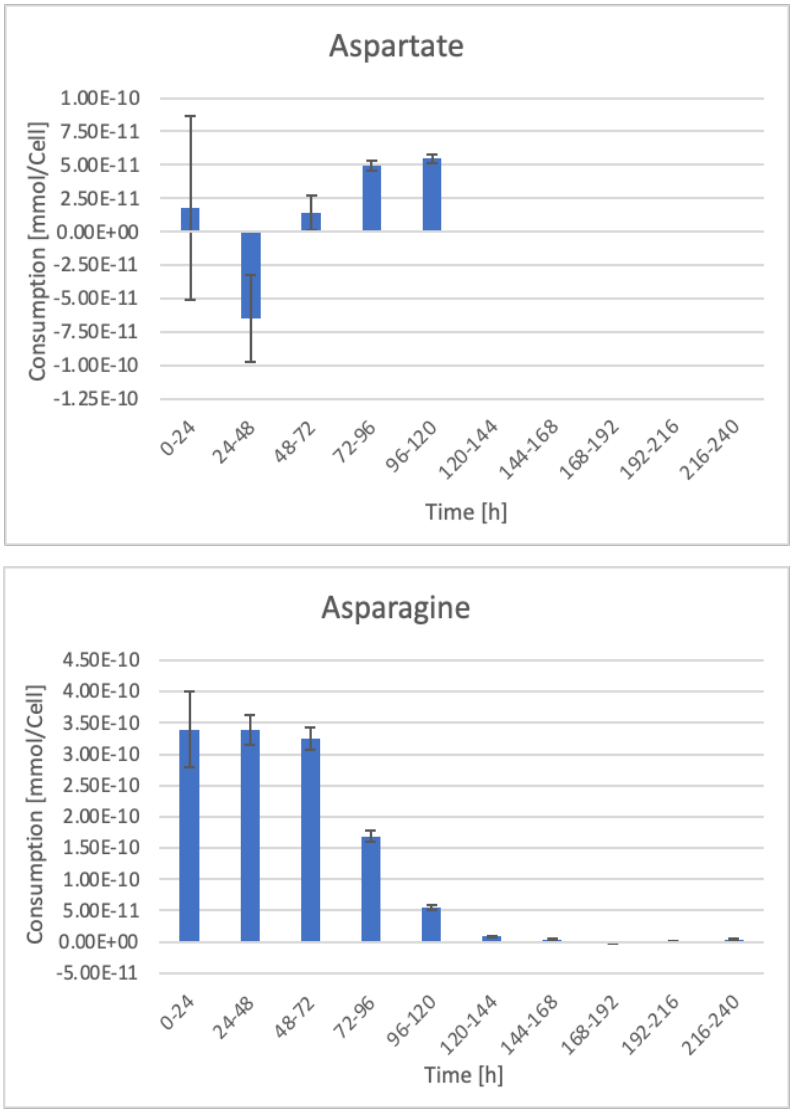


Figure 5.12: Specific Consumption of Aspartate and Asparagine

The sudden consumption of pantothenic acid and pyridoxine between 192 and 216 hours, which was noted previously, is emphasized more greatly here. Interestingly, this instantaneous consumption is not sustained after the 24-hour interval that it is observed. Similarly, biotin exhibited high specific consumption that was only observed between 24 and 48 hours, followed by a rapid decrease in specific consumption. The specific consumption profiles of all three of these vitamins is shown in Figure 5.13.

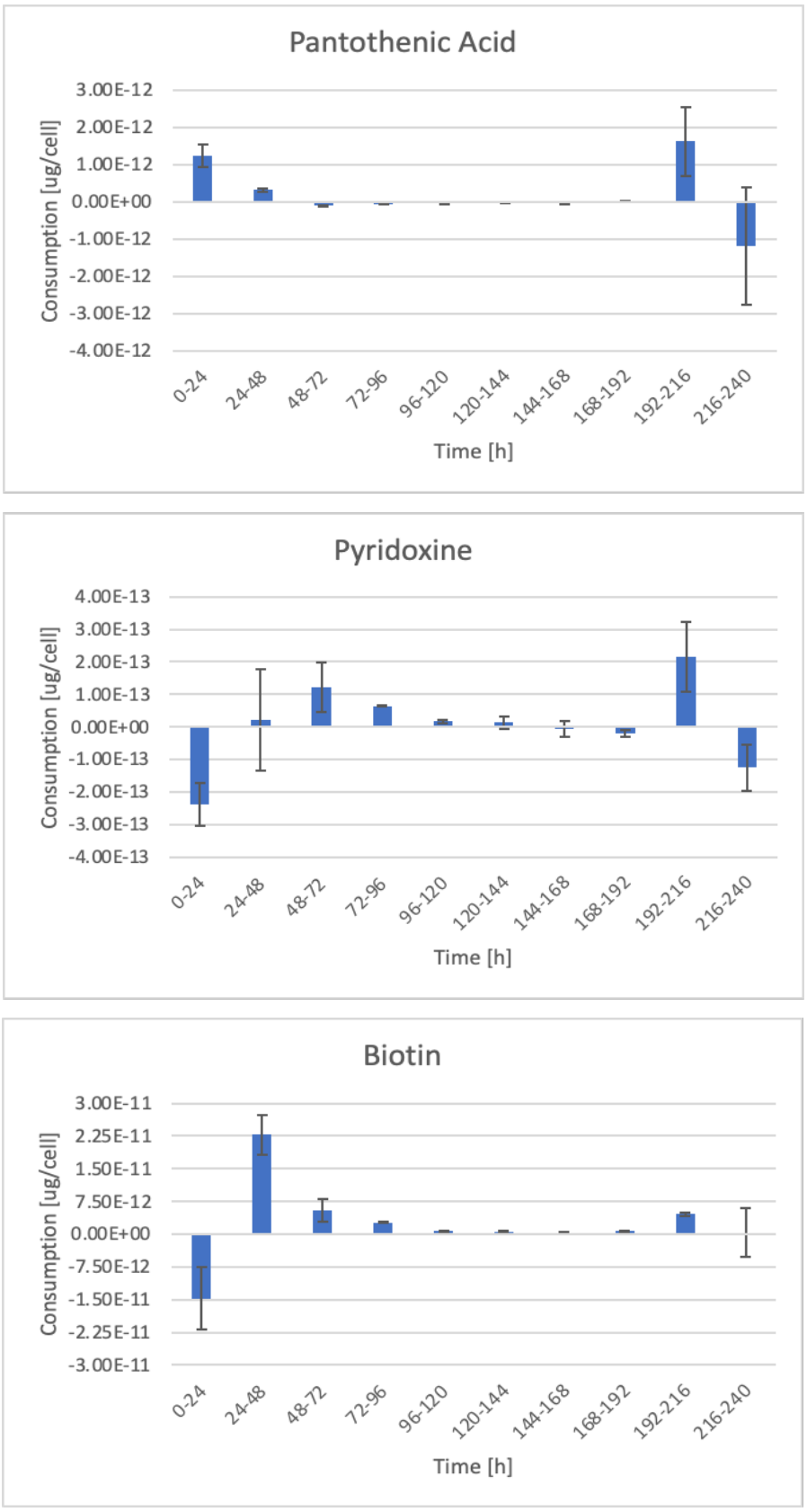


Figure 5.13: Specific Consumption of Pantothenic Acid, Pyridoxine and Biotin

5.5 Discussion

5.5.1 Glycine Production and the Alanine Shift

Glycine production is hypothesized to be a result of serine hydroxymethyltransferase (SHMT) activity, an enzyme which catalyzes the interconversion of serine (which was steadily consumed) into glycine. The precise reason for this activity is yet to be fully understood, but both amino acids are known to be involved in nucleotide biosynthesis, so it is possible that this GS-CHO clone has a preferential need for glycine over serine [67]. The alanine shift is likely a result of the metabolic shifts surrounding energy metabolism; alanine can be used for the direct synthesis of pyruvate, which the cells may resort to in the instance that glucose has been substantially depleted and glycolysis is no longer a reliable source of pyruvate production [62]. Of the amino acids depicted in Figure 5.3, which clustered similarly in the HCA based on their production patterns, the metabolic shift for alanine was concluded to be the most interesting and significant. Its strong role in energy metabolism led to the investigation of other amino acids or vitamins with possibly related significant shifts.

5.5.2 Glutamine Shift and Glutamate Exhaustion

The glutamine shift from production to consumption at 120 hours occurs just 25 hours prior to the Alanine shift and can be explained by one of the key anaplerotic reactions of the TCA cycle which involves the biosynthesis of glutamate from glutamine for the ultimate conversion into α-ketoglutarate, a TCA cycle intermediate; once the cells have exhausted their free supply of extracellular glutamate, it must be endogenously produced to maintain the TCA cycle [68]. The exhaustion of glutamine at 144 hours coincides with the alanine

shift, which is sensible because of the roles they play in energy metabolism; the exhaustion of both glutamine and glutamate severely limits α-ketoglutarate as a viable entry point into the TCA cycle, so the cells shift towards alanine consumption for the purpose of pyruvate generation. The temporary halt in ammonia production between 96 and 120 hours is possibly a result of ammonia being a byproduct of the transaminase reactions for glutamine and alanine conversion into glutamate and pyruvate respectfully.

5.5.3 Asparagine and Aspartate Exhaustion

Asparagine and aspartate are two other amino acids that were observed to be completely exhausted. Aspartate is a critical nutrient because it is the only known free amino acid that can be used for the anaplerotic synthesis of oxaloacetate, the final product of the TCA cycle, making it functionally comparable to the relationship between glutamate and ketoglutarate [66]. However, while glutamate can be synthesized from numerous other free amino acids, aspartate can only be synthesized from asparagine, which was observed to be exhausted at 120 hours simultaneously with aspartate. This too was a likely trigger for the shift towards alanine consumption at 144 hours, because at that point, the oxaloacetate could only be produced from malate within the TCA cycle, or from pyruvate directly (catalyzed by pyruvate carboxylase) [66].

5.5.4 Supernatant Concentrations of Pantothenic Acid, Pyridoxal and Pyridoxine

Pantothenic acid is required for the synthesis of coenzyme A (CoA), which plays a key role in both lipid metabolism and the TCA cycle, with the formation of acetyl-CoA and succinyl-CoA [69]. Referring back to the graph of viable cell density in Figure 4.1, it can be

seen that the number of viable cells was in rapid decline at 216 hours, and it must also be noted that glucose concentration was almost entirely depleted just 24 hours prior to this. Therefore, it is hypothesized that the complete exhaustion of extracellular glucose prompted a stark shift in cellular dependency on the TCA cycle for energy production, and therefore suddenly increased cellular demand for pantothenic acid. Additionally, the concentration of pyridoxine also yielded a somewhat sudden decrease at 216 hours, albeit to a lesser extent. Pyridoxine is an important cofactor in many catabolic and biosynthetic reactions for amino acid metabolism, making it similarly important for energy generation and related to TCA cycle upregulation; it is also capable of neutralizing ROS, so its relationship to TCA cycle activity and redox stress is clear [63]. The exhaustion of pyridoxal around 72 hours is sensible, because pyridoxine has been reported to be the most biologically active member of the B_6 family, explaining initial cellular preference for it over pyridoxal [69].

5.5.5 Relating the PCA to the HCA

It is not surprising that a majority of the variation can be captured by a single principal component in the PCA because by looking at the HCA heatmap in Figure 5.2, it can clearly be seen that a majority of the variables exhibit significantly similar concentration profiles, characterized largely by high concentrations during the lag and exponential phases, followed by lower concentrations and consumption during the stationary phase. Therefore, the same clusters that were identified in the HCA were likewise seen in the PCA.

5.5.6 Calculation of Specific Consumption and Production

The specific consumption and production profiles were in many cases quite different from the patterns observed in the extracellular concentration profiles. While it may initially seem as if these calculations would yield similar results to the extracellular concentration profiles alone, the reality is that it can actually be quite transformative. Much like calculating a derivative, it shifts the focus from the magnitude of the values themselves to the magnitude of change between them. Additionally, the normalization to viable cell number at a specific point in time elucidates metabolic activity that may not be clear during time points of lower cell density, such as the lag phase. A prime example of this is the specific production of mAb (Figure 4.2) – the graph of extracellular mAb concentration suggests that mAb production is low during the lag phase, but the graph of specific production communicates the opposite. When accounting for the lower cell density, it can be seen that productivity was consistently at its highest during the first four days of the culture and decreased in steady increments until reaching its low at 216 hours, the first observation of rapid cell death. This is contradictory to some works that have been published, which show that upregulated TCA cycle activity associated with the stationary phase of CHO cultures correlated positively with productivity in fed-batch cultures [70]. This was hypothesized to be a result of the comparatively efficient energy metabolism of the TCA cycle and lactate consumption compared to high glycolytic flux and lactate production. The likely causation for the differences observed between the data shared here and published findings is the application of fed-batch culture instead of batch. The steady feeding of key nutrients, such as glutamate (which was comparatively exhausted near the beginning of the stationary phase in the cultures described in this thesis) allows for more efficient metabolic activity that is less reliant

on anaplerotic pathways which accumulate ammonia, increasing the toxicity of the environment and limiting the free amino acid availability for mAb construction. A greater availability of key nutrients for energy metabolism allows for the employment of more free amino acids for mAb construction rather than biosynthesis of depleted amino acids, translating to enhanced specific productivity and culture longevity and ultimately increasing mAb titer. This is a key advantage of fed-batch over batch culture, as described in the first chapter, and is the motivation behind the exclusive employment of fed-batch methodology in much of the current literature encompassing characterization and optimization of CHO-based biologics production.

5.5.7 Specific Consumption of Pantothenic Acid and Biotin

Further research will need to be conducted in order to confidently suggest probable cause for the high specific consumption of pantothenic acid during the first 24 hours, as there is no clear explanation for this observation in the data. A clear shift in the consumption of biotin is observed during the first 48 hours before achieving steady-state – a complete lack of consumption is observed during the lag phase, followed by high consumption during the first 24 hours of exponential growth. Biotin, or vitamin B_7, is an essential cofactor for multiple carboxylase enzymes, such as pyruvate and acetyl-CoA carboxylase, which play important roles in fatty acid synthesis [71]. Given that fatty acids are a key component of phospholipid membranes, it can be hypothesized that biotin is needed for the synthesis of lipid biomass for membrane construction during cell proliferation, which would explain its high necessity during exponential growth but not the lag phase. This, however, does not explain the rapid decrease in biotin demand after 48 hours, given that exponential growth

is sustained until 96 hours. Further optimization of the vitamins quantification protocol will need to be performed in order to be confident in these results, given the poor shape and resolution of many of the HPLC chromatographs. Overall, however, the ANOVA-supported multivariate analysis proved to reduce insignificant noise in the heatmap while still identifying all the same key shifts and clusters from the previous analysis.

CHAPTER 6

CONCLUSION AND NEXT STEPS

6.1 Summary and Conclusions

Efficient sample collection and proper experimental planning for cell culture allows for a multifaceted approach to data analysis to ultimately obtain a more complete characterization. Quantitative methods for the determination of extracellular amino acids, vitamins and metabolites are critical to profiling the metabolic behavior of the culture and optimizing experimental parameters, such as culture conditions and media composition. However, it can be difficult to discern significant findings and correlations between variables in larger amounts of data without a standardized and streamlined approach for analysis.

In this work, GS-CHO cells were grown in batch culture, and samples were collected daily for measurement of viability, ATP, mAb titer, and extracellular amino acids, vitamins and common metabolites. The creation of a clean and robust Excel template file that was specifically designed to accommodate the formatting of electronic reports generated by our lab equipment facilitated easy data consolidation and standardized initial data processing and visualization. Programming in RStudio, with the help of existing free online resources, was used to standardize variables of differing units and magnitudes by computing their z-score and provided a toolset to perform and visualize PCA and HCA.

The four stages of culture growth (lag, exponential, stationary, and decline) exhibit distinct metabolic characteristics, which were identified by the unsupervised clustering of extracellular metabolite profiles in both HCA and PCA. While the consumption and production patterns of many nutrients and metabolites were deemed insignificant to the decline

of viability or mAb titer, numerous metabolic shifts with known significance were elucidated, as well as a handful that may require further exploration to fully understand.

It is hypothesized that the exhaustion of glutamate, aspartate and asparagine after 120 hours yielded significant metabolic consequences by limiting the synthesis of TCA cycle intermediates ketoglutarate and oxaloacetate, ultimately resulting in increased dependence on ammonia-producing anaplerotic pathways, pyruvate synthesis from alanine, and generalized redox stress from upregulation of the TCA cycle. High specific consumption of glucose during the first 72 hours of culture, which coincided precisely with the production of lactate, validated published claims that the metabolic shift away from lactic acid fermentation is not associated with glucose depletion, but rather a significant decrease in the rate of specific glucose consumption [62]. Understanding and characterizing the lactate shift in recombinant antibody-producing mammalian cells has yet to be fully complete but is essential for the optimization of productivity.

6.2 Future Steps

The use of these findings to inform basal media supplementation and a feeding strategy for fed-batch culture is the necessary next step to optimize mAb production. It is clear that at a minimum, glutamate and aspartate need to be fed to the culture prior to their exhaustion during early stationary phase. It is also worth exploring the reasons for the high specific consumption of many vitamins during lag and early exponential (B_1, C, B_2, B_3, B_5, Pyridoxal, B_{12}) while others exhibited suddenly high specific consumption at the start of viability decline (B_5, Pyridoxine). Future experiments will entail the employment of a Sartorius Stedim single-use benchtop bioreactor for the fed-batch culturing of GS-CHO cells for

mAb production. The MVA techniques utilized in this work will be used to inform and validate a mathematical model designed by a postdoctoral researcher in the lab for the iterative optimization of media composition and feeding strategy for maximizing culture longevity and volumetric productivity.

APPENDIX

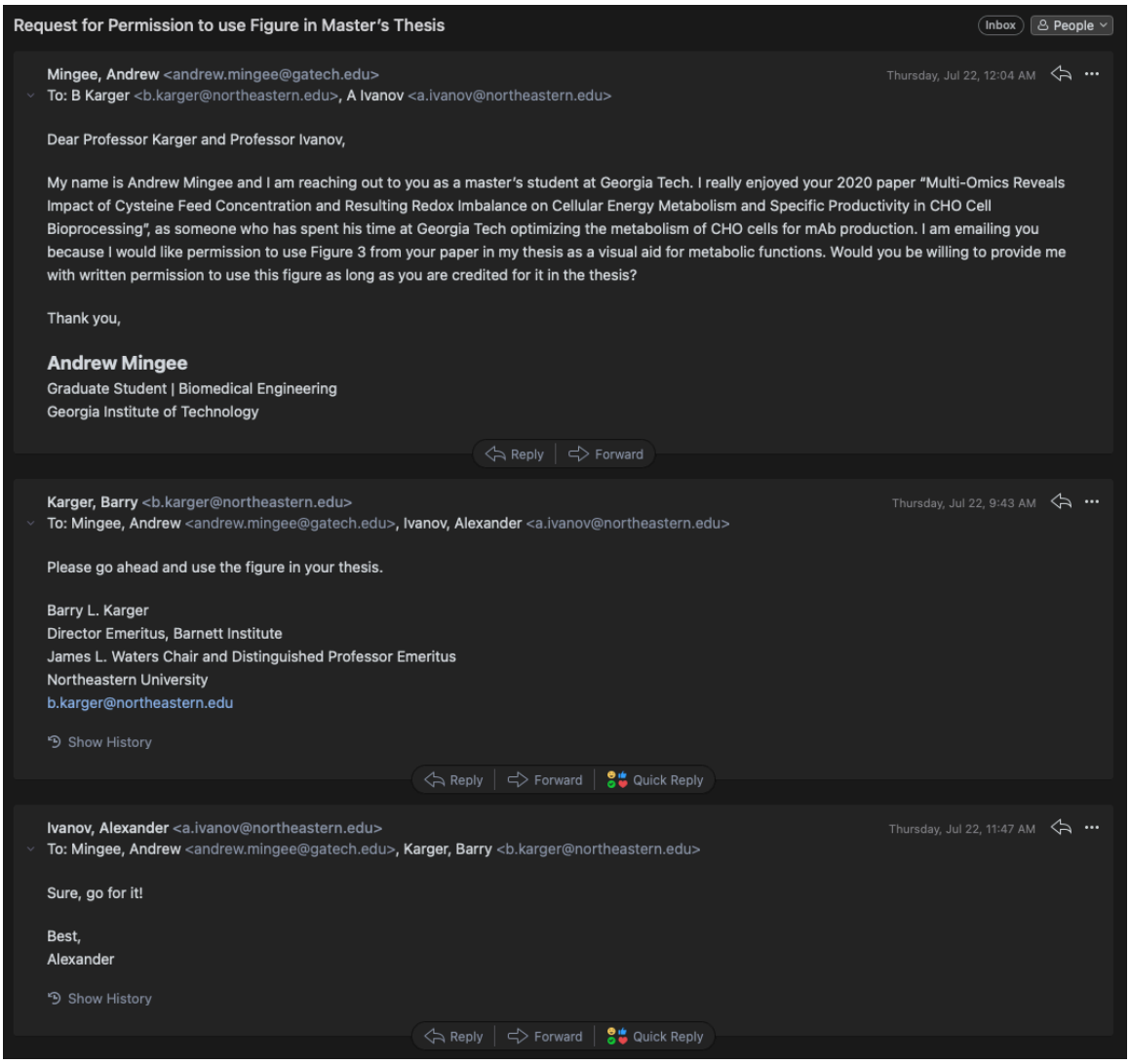


Figure A.1: Written Permission to Use Copyrighted Figure 5.1

REFERENCES

[1] A. Kumar, I. A. Udugama, C. L. Gargalo, and K. V. Gernaey, "Why is batch processing still dominating the biologics landscape? Towards an integrated continuous bioprocessing alternative," *Processes*, vol. 8, no. 12, pp. 1–19, 2020, doi: 10.3390/pr8121641.

[2] Pfizer *et al.*, *A-Mab: A Case Study in Bioprocess Development*, no. October. 2009. doi: 10.1016/j.jiec.2018.01.030.

[3] V. Kumar, A. Bhalla, and A. S. Rathore, "Design of experiments applications in bioprocessing: Concepts and approach," *Biotechnol. Prog.*, vol. 30, no. 1, pp. 86–99, 2014, doi: 10.1002/btpr.1821.

[4] E. Daskalaki, N. J. Pillon, A. Krook, C. E. Wheelock, and A. Checa, "The influence of culture media upon observed cell secretome metabolite profiles: The balance between cell viability and data interpretability," *Anal. Chim. Acta*, vol. 1037, pp. 338–350, 2018, doi: 10.1016/j.aca.2018.04.034.

[5] A. L. Grilo and A. Mantalaris, "The Increasingly Human and Profitable Monoclonal Antibody Market," *Trends Biotechnol.*, vol. xx, pp. 1–7, 2014, doi: 10.1016/j.tibtech.2018.05.014.

[6] D. Shek, S. A. Read, G. Ahlenstiel, and I. Piatkov, "Pharmacogenetics of anticancer monoclonal antibodies," *Cancer Drug Resist.*, pp. 69–81, 2019, doi: 10.20517/cdr.2018.20.

[7] H. W. Schroeder and L. Cavacini, "Structure and function of immunoglobulins," *J. Allergy Clin. Immunol.*, vol. 125, no. 2, pp. S41–S52, Feb. 2010, doi: 10.1016/j.jaci.2009.09.046.

[8] B. Moritz and J. O. Stracke, "Assessment of disulfide and hinge modifications in monoclonal antibodies: General," *ELECTROPHORESIS*, vol. 38, no. 6, pp. 769–785, Mar. 2017, doi: 10.1002/elps.201600425.

[9] Absolute Antibody, "Antibody Fragments," *Absolute Antibody*. https://absoluteantibody.com/antibody-resources/antibody-engineering/antibody-fragments/ (accessed May 31, 2021).

[10] E. A. van Erp, W. Luytjes, G. Ferwerda, and P. B. van Kasteren, "Fc-Mediated Antibody Effector Functions During Respiratory Syncytial Virus Infection and Disease," *Front. Immunol.*, vol. 10, 2019, doi: 10.3389/fimmu.2019.00548.

[11] R.-M. Lu *et al.*, "Development of therapeutic antibodies for the treatment of diseases," *J. Biomed. Sci.*, vol. 27, no. 1, p. 1, Dec. 2020, doi: 10.1186/s12929-019-0592-z.

[12] G. L. Ada and S. G. Nossal, "The Clonal-Selection Theory," *Sci. Am.*, vol. 257, no. 2, pp. 62–69, Aug. 1987, doi: 10.1038/scientificamerican0887-62.

[13] D. Fulcher and A. Basten, "B cell life span: A review," *Immunol. Cell Biol.*, vol. 75, no. 5, pp. 446–455, Oct. 1997, doi: 10.1038/icb.1997.69.

[14] H. A. Parray *et al.*, "Hybridoma technology a versatile method for isolation of monoclonal antibodies, its applicability across species, limitations, advancement and

future perspectives," *Int. Immunopharmacol.*, vol. 85, p. 106639, Aug. 2020, doi: 10.1016/j.intimp.2020.106639.

[15] Millipore Sigma, "Hybridoma Cell Technology," *Hybridoma Fusion Partners Cell Lines*. https://www.sigmaaldrich.com/technical-documents/protocols/biology/cell-culture/hybridoma-fusion.html (accessed May 31, 2021).

[16] P. E. Rao and D. J. Kroon, "Orthoclone OKT3," in *Stability and Characterization of Protein and Peptide Drugs: Case Histories*, Y. J. Wang and R. Pearlman, Eds. Boston, MA: Springer US, 1993, pp. 135–158. doi: 10.1007/978-1-4899-1236-7_4.

[17] L. Ledsgaard, M. Kilstrup, A. Karatt-Vellatt, J. McCafferty, and A. Laustsen, "Basics of Antibody Phage Display Technology," *Toxins*, vol. 10, no. 6, p. 236, Jun. 2018, doi: 10.3390/toxins10060236.

[18] "Sir Greg Winter wins the 2018 Nobel Prize in Chemistry," *University of Cambridge*, Oct. 03, 2018. https://www.cam.ac.uk/research/news/sir-greg-winter-wins-the-2018-nobel-prize-in-chemistry (accessed May 26, 2021).

[19] J. Boonyaratanakornkit and J. J. Taylor, "Techniques to Study Antigen-Specific B Cell Responses," *Front. Immunol.*, vol. 10, p. 1694, Jul. 2019, doi: 10.3389/fimmu.2019.01694.

[20] J. L. Spidel, B. Vaessen, Y. Y. Chan, L. Grasso, and J. B. Kline, "Rapid high-throughput cloning and stable expression of antibodies in HEK293 cells," *J. Immunol. Methods*, vol. 439, pp. 50–58, Dec. 2016, doi: 10.1016/j.jim.2016.09.007.

[21] E. V. Voronina, Y. A. Seregin, N. A. Litvinova, V. I. Shvets, and R. R. Shukurov, "Design of a stable cell line producing a recombinant monoclonal anti-TNFα antibody based on a CHO cell line," *SpringerPlus*, vol. 5, no. 1, p. 1584, Dec. 2016, doi: 10.1186/s40064-016-3213-2.

[22] F. Li, N. Vijayasankaran, A. (Yijuan) Shen, R. Kiss, and A. Amanullah, "Cell culture processes for monoclonal antibody production," *mAbs*, vol. 2, no. 5, pp. 466–479, Sep. 2010, doi: 10.4161/mabs.2.5.12720.

[23] A. Lima Grilo, "Integrated multi-scale models for biologics process development," Imperial College London, 2019.

[24] J. Dumont, D. Euwart, B. Mei, S. Estes, and R. Kshirsagar, "Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives," *Crit. Rev. Biotechnol.*, vol. 36, no. 6, pp. 1110–1122, 2016, doi: 10.3109/07388551.2015.1084266.

[25] S. Fischer, R. Handrick, and K. Otte, "The art of CHO cell engineering: A comprehensive retrospect and future perspectives," *Biotechnol. Adv.*, vol. 33, no. 8, pp. 1878–1896, 2015, doi: 10.1016/j.biotechadv.2015.10.015.

[26] F. Zhang, X. Sun, X. Yi, and Y. Zhang, "Metabolic characteristics of recombinant Chinese hamster ovary cells expressing glutamine synthetase in presence and absence of glutamine," *Cytotechnology*, vol. 51, no. 1, pp. 21–28, May 2006, doi: 10.1007/s10616-006-9010-y.

[27] P.-C. Lin *et al.*, "Attenuated glutamine synthetase as a selection marker in CHO cells to efficiently isolate highly productive stable cells for the production of antibodies and other biologics," *mAbs*, vol. 11, no. 5, pp. 965–976, Jul. 2019, doi: 10.1080/19420862.2019.1612690.

[28] R. E. Kingston, R. J. Kaufman, C. R. Bebbington, and M. R. Rolfe, “Amplification Using Cell Expression Vectors,” *Curr. Protoc. Mol. Biol.*, vol. 60, no. 1, Oct. 2002, doi: 10.1002/0471142727.mb1623s60.

[29] T. Lai, Y. Yang, and S. Ng, “Advances in Mammalian Cell Line Development Technologies for Recombinant Protein Production,” *Pharmaceuticals*, vol. 6, no. 5, pp. 579–603, Apr. 2013, doi: 10.3390/ph6050579.

[30] A. A. Shukla and J. Thömmes, “Recent advances in large-scale production of monoclonal antibodies and related proteins,” *Trends Biotechnol.*, vol. 28, no. 5, pp. 253–261, 2010, doi: 10.1016/j.tibtech.2010.02.001.

[31] M. M. Zhu, M. Mollet, R. S. Hubert, Y. S. Kyung, and G. G. Zhang, “Industrial Production of Therapeutic Proteins: Cell Lines, Cell Culture, and Purification,” in *Handbook of Industrial Chemistry and Biotechnology*, J. A. Kent, T. V. Bommaraju, and S. D. Barnicki, Eds. Cham: Springer International Publishing, 2017, pp. 1639–1669. doi: 10.1007/978-3-319-52287-6_29.

[32] F. LeFloch *et al.*, “Related effects of cell adaptation to serum-free conditions on murine EPO production and glycosylation by CHO cells,” *Cytotechnology*, vol. 52, no. 1, pp. 39–53, Sep. 2006, doi: 10.1007/s10616-006-9039-y.

[33] A. L. Grilo and A. Mantalaris, “A Predictive Mathematical Model of Cell Cycle, Metabolism, and Apoptosis of Monoclonal Antibody-Producing GS–NS0 Cells,” *Biotechnol. J.*, vol. 14, no. 11, p. 1800573, 2019, doi: https://doi.org/10.1002/biot.201800573.

[34] A. L. Quiroga, “Mathematical modelling and experimental validation for optimisation and control of mammalian cell culture systems,” Imperial College London, 2017.

[35] A. A. Shukla and U. Gottschalk, “Single-use disposable technologies for biopharmaceutical manufacturing,” *Trends Biotechnol.*, vol. 31, no. 3, pp. 147–154, Mar. 2013, doi: 10.1016/j.tibtech.2012.10.004.

[36] Eppendorf, “Perfusion - Making the Most out of Your Working Volume,” *Eppendorf: Handling Solutions*, 2017. https://handling-solutions.eppendorf.com/cell-handling/bioprocess/processes-and-applications/detailview/news/perfusion-making-the-most-out-of-your-working-volume/

[37] N. Zamboni, A. Saghatelian, and G. J. Patti, “Defining the Metabolome: Size, Flux, and Regulation,” *Mol. Cell*, vol. 58, no. 4, pp. 699–706, May 2015, doi: 10.1016/j.molcel.2015.04.021.

[38] V. G. Dhara, H. M. Naik, N. I. Majewska, and M. J. Betenbaugh, “Recombinant Antibody Production in CHO and NS0 Cells: Differences and Similarities,” *BioDrugs*, vol. 32, pp. 571–584, 2018, doi: 10.1007/s40259-018-0319-9.

[39] P. R. Graves and T. A. J. Haystead, “Molecular Biologist’s Guide to Proteomics,” *Microbiol. Mol. Biol. Rev.*, vol. 66, no. 1, pp. 39–63, Mar. 2002, doi: 10.1128/MMBR.66.1.39-63.2002.

[40] B. G. Kremkow, J. Y. Baik, M. L. MacDonald, and K. H. Lee, “CHOgenome.org 2.0: Genome resources and website updates.,” *Biotechnol. J.*, pp. 931–938, 2015.

[41] M. A. Valle, M. B. Kester, A. L. Burns, S. J. Marx, A. M. Spiegel, and J. Shiloach, “Production and purification of human menin from Drosophila melanogaster S2 cells

using stirred tank reactor," *Cytotechnology*, vol. 35, no. 2, pp. 127–135, Mar. 2001, doi: 10.1023/A:1017586523710.
[42] T. B. Baker *et al.*, "Implementing Clinical Research Using Factorial Designs: A Primer," *Behav. Ther.*, vol. 48, no. 4, pp. 567–580, Jul. 2017, doi: 10.1016/j.beth.2016.12.005.
[43] S. Wold, K. Esbensen, and P. Geladi, "Principal Component Analysis," p. 16.
[44] G. W. Milligan and M. C. Cooper, "A study of standardization of variables in cluster analysis," *J. Classif.*, vol. 5, no. 2, pp. 181–204, Sep. 1988, doi: 10.1007/BF01897163.
[45] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.
[46] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999, doi: 10.1145/331499.331504.
[47] J. Tabak, *Geometry: The Language of Space and Form*. Infobase Publishing, 2014.
[48] A. Jarman, "Hierarchical Cluster Analysis: Comparison of Single linkage,Complete linkage, Average linkage and Centroid Linkage Method," Feb. 2020. doi: 10.13140/RG.2.2.11388.90240.
[49] B. Dealmakers, "Moving up with the monoclonals," *Biopharma Deal.*, Sep. 2019, doi: 10.1038/d43747-020-00765-2.
[50] A. H. Sharpe and K. E. Pauken, "The diverse functions of the PD1 inhibitory pathway," *Nat. Rev. Immunol.*, vol. 18, no. 3, Art. no. 3, Mar. 2018, doi: 10.1038/nri.2017.108.
[51] M. Yasunaga, "Antibody therapeutics and immunoregulation in cancer and autoimmune disease," *Semin. Cancer Biol.*, vol. 64, pp. 1–12, Aug. 2020, doi: 10.1016/j.semcancer.2019.06.001.
[52] P. J. Mease, "Adalimumab in the treatment of arthritis," *Ther. Clin. Risk Manag.*, vol. 3, no. 1, pp. 133–148, Mar. 2007.
[53] D. C. Julien, S. Behnke, G. Wang, G. K. Murdoch, and R. A. Hill, "Utilization of monoclonal antibody-targeted nanomaterials in the treatment of cancer," *mAbs*, vol. 3, no. 5, pp. 467–478, 2011, doi: 10.4161/mabs.3.5.16089.
[54] D. Howard *et al.*, "Antibody–drug conjugates and other nanomedicines: the frontier of gynaecological cancer treatment," *Interface Focus*, vol. 6, no. 6, Dec. 2016, doi: 10.1098/rsfs.2016.0054.
[55] A. Juan, F. J. Cimas, I. Bravo, A. Pandiella, A. Ocaña, and C. Alonso-Moreno, "An Overview of Antibody Conjugated Polymeric Nanoparticles for Breast Cancer Therapy," *Pharmaceutics*, vol. 12, no. 9, Aug. 2020, doi: 10.3390/pharmaceutics12090802.
[56] Thermo Fisher Scientific, "Overview of ELISA." //www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html (accessed May 27, 2021).
[57] W. Ansar and S. Ghosh, "Monoclonal Antibodies: A Tool in Clinical Research," *Indian J. Clin. Med.*, vol. 4, p. IJCM.S11968, Jan. 2013, doi: 10.4137/IJCM.S11968.

[58] A. L. Quiroga-Campano, N. Panoskaltsis, and A. Mantalaris, “Energy-based culture medium design for biomanufacturing optimization: A case study in monoclonal antibody production by GS-NS0 cells,” *Metab. Eng.*, vol. 47, pp. 21–30, May 2018, doi: 10.1016/j.ymben.2018.02.013.
[59] Thermo Fisher Scientific, “MAbPac Protein A Column Product Manual.” Thermo Fisher Scientific, 2013.
[60] Henderson Jr, John W, Brooks, Anne, and Agilent Technologies, Inc., “Improved Amino Acid Methods using Agilent ZORBAX Eclipse Plus C18 Columns for a Variety of Agilent LC Instrumentation and Separation Goals.” Agilent Technologies, Inc., Apr. 06, 2010.
[61] Thermo Fisher Scientific, “Determination of Water- and Fat-Soluble Vitamins by HPLC.” 2017.
[62] B. C. Mulukutla, M. Gramer, and W.-S. Hu, “On metabolic shift to lactate consumption in fed-batch culture of mammalian cells,” *Metab. Eng.*, vol. 14, no. 2, pp. 138–149, Mar. 2012, doi: 10.1016/j.ymben.2011.12.006.
[63] M. Parra, S. Stahl, and H. Hellmann, “Vitamin B6 and Its Role in Cell Metabolism and Physiology,” *Cells*, vol. 7, no. 7, Jul. 2018, doi: 10.3390/cells7070084.
[64] Y. Bai, C. Wu, and W. L. W. Ling, “Role of Iron and Sodium Citrate in Animal Protein-Free CHO Cell Culture Medium on Cell Growth and Monoclonal Antibody Production,” *Biotechnol. Prog.*, pp. 209–219, 2011, doi: 10.1002/btpr.513.
[65] A. S. Ali *et al.*, “Multi-Omics Reveals Impact of Cysteine Feed Concentration and Resulting Redox Imbalance on Cellular Energy Metabolism and Specific Productivity in CHO Cell Bioprocessing,” *Biotechnol. J.*, vol. 15, no. 8, p. 1900565, 2020, doi: https://doi.org/10.1002/biot.201900565.
[66] R. Perumcheril, *Self Assessment And Review Of Biochemistry*, 6th ed. Jaypee Brothers Medical Publishers, 2020.
[67] M. R. Narkewicz, S. D. Sauls, S. S. Tjoa, C. Teng, and P. V. Fennessey, “Evidence for intracellular partitioning of serine and glycine metabolism in Chinese hamster ovary cells.,” *Biochem. J.*, vol. 313, no. Pt 3, pp. 991–996, Feb. 1996.
[68] D. A. Bender, “TRICARBOXYLIC ACID CYCLE,” in *Encyclopedia of Food Sciences and Nutrition (Second Edition)*, B. Caballero, Ed. Oxford: Academic Press, 2003, pp. 5851–5856. doi: 10.1016/B0-12-227055-X/01363-8.
[69] H. M. Said, “Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency,” in *Encyclopedia of Gastroenterology*, L. R. Johnson, Ed. New York: Elsevier, 2004, pp. 631–637. doi: 10.1016/B0-12-386860-2/00096-4.
[70] X. Pan, C. Dalm, R. H. Wijffels, and D. E. Martens, “Metabolic characterization of a CHO cell size increase phase in fed-batch cultures,” pp. 8101–8113, 2017, doi: 10.1007/s00253-017-8531-y.
[71] “Biotin in Cell Culture,” *Sigma-Aldrich*. https://www.sigmaaldrich.com/life-science/cell-culture/learning-center/media-expert/biotin.html (accessed Apr. 23, 2021).